(51) International Patent Classification 6 :	A2	(11) International Publication Number:	WO 99/19492
C12N 15/29, C07K 14/415, C12N 5/04, 5/14, A01H 5/00, C12N 15/10, 15/82		(43) International Publication Date:	22 April 1999 (22.04.99)

(21) International Application Number: PCT/EP98/06977

(22) International Filing Date: 9 October 1998 (09.10.98)

(30) Priority Data:
PO 9745   10 October 1997 (10.10.97)   AU

(71) Applicant (for all designated States except US): RHONE-POULENC AGRO [FR/FR]; 14/20, rue Pierre Baizet, F-69009 Lyon (FR).

(72) Inventors; and
(75) Inventors/Applicants (for US only): DOUTRIAUX, Marie-Pascale [FR/FR]; 64, route de Villebon, F-91160 Saulx les Chartreux (FR). BETZNER, Andreas, Stefan [AU/AU]; 40 Dallachy Place, Page, ACT 2614 (AU). FREYSSINET, Georges [FR/FR]; 21, rue de Nervieux, F-69450 Saint Cyr au Mont d'Or (FR). PEREZ, Pascal [FR/FR]; 17, chemin de la Pradelle, Varennes, F-63450 Chanonat (FR).

(74) Agent: GENIN, Patrick; Rhône-Poulenc Agro, DPI, 14/20, rue Pierre Baizet, F-69009 Lyon (FR).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHODS FOR OBTAINING PLANT VARIETIES

(57) Abstract

An isolated and purified DNA molecule comprising a polynucleotide sequence encoding a polypeptide functionally involved in the DNA mismatch repair system of a plant.

Methods for Obtaining Plant Varieties

TECHNICAL FIELD

The present invention relates to nucleotide sequences which encode polypeptides
involved in the DNA mismatch repair systems of plants, and to the polypeptides encoded
by those nucleotide sequences. The invention also relates to nucleotide sequences and
polypeptide sequences for use in altering the DNA mismatch repair system in plants. The
invention also relates to a process for altering the DNA mismatch repair system of a plant
cell, to a process for increasing genetic variations in plants and to processes for obtaining
plants having a desired characteristic.

BACKGROUND OF THE INVENTION

Plant breeding essentially relies on and makes use of genetic variation which occurs
naturally within and between members of a family, a genus, a species or a subspecies.
Another source of genetic variation is the introduction of genes from other organisms
which may or may not be related to the host plant.

Allelic loci or non-allelic genes which constitute or contribute to desired quantitative
(e.g. growth performance, yield, etc.) or qualitative (e.g. deposition, content and
composition of seed storage products; pathogen resistance genes; etc.) traits that are
absent, incomplete or inefficient in a species or subspecies of interest are typically
introduced by the plant breeder from other species or subspecies, or *de novo*. This
introduction is often done by crossing, provided that the species to be crossed are sexually
compatible. Other means of introducing genomes, individual chromosomes or genes into
plant cells or plants are well known in the art. They include cell fusion, chemically aided
transfection (Schocher et al., 1986, Biotechnology 4: 1093) and ballistic (McCabe et al.,
1988, Biotechnology 6: 923), microinjection (Neuhaus et al., 1987, TAG 75: 30),
electroporation of protoplasts (Chupeau et al., 1989, Biotechnology 7: 53) or microbial
transformation methods such as Agrobacterium mediated transformation (Horsch et al.,
1985, Science 227: 1229; Hiei et al., 1996, Biotechnology 14: 745).

However, when a foreign genome, chromosome or gene is introduced into a plant, it
will often segregate in subsequent generations from the genome of the recipient plant or
plant cell during mitotic and meiotic cell divisions and, in consequence, become lost from
the host plant or plant cell into which it had been introduced. Occasionally, however, the
introduced genome, chromosome or gene physically combines entirely or in part with the
genome, chromosome or gene of the host plant or plant cell in a process which is called
recombination.

Recombination involves the exchange of covalent linkages between DNA molecules
in regions of identical or similar sequence. It is referred to here as homologous
recombination if donor and recipient DNA are identical or nearly identical (at least 99%

base sequence identity), and as homeologous recombination if donor and recipient DNA are not identical but are similar (less than 99% base sequence identity).

The ability of two genomes, chromosomes or genes to recombine is known to depend largely on the evolutionary relation between them and thus on the degree of
5 sequence similarity between the two DNA molecules. Whereas homologous recombination is frequently observed during mitosis and meiosis, homeologous recombination is rarely or never seen.

From a breeder's perspective, the limits within which homologous recombination occurs, therefore, define a genetic barrier between species, varieties or lines, in contrast
10 to homeologous recombination which can break this barrier. Homeologous recombination is thus of great importance for plant breeding. Accordingly there is a need for a process for enhancing the frequency of homeologous recombination in plants. In particular, there is a need for a process of increasing homeologous recombination to significantly shorten the length of breeding programs by reducing the number of crosses required to obtain an
15 otherwise rare recombination event.

At least in *Escherichia coli*, homologous and homeologous recombination are known to share a common pathway that requires among others the proteins RecA, RecB, RecC, RecD and makes use of the SOS induced RuvA and RuvB, respectively. It has been suggested that mating induced recombination follows the Double-Strand Break Repair
20 model (Szostak et al., 1983, Cell 33, 25-35), which is widely used to describe genetic recombination in eukaryotes. Following the alignment of homologous or homeologous DNA double helices the RecA protein mediates an exchange of a single DNA strand from the donor helix to the aligned recipient DNA helix. The incoming strand screens the recipient helix for sequence complementarity, seeking to form a heteroduplex by hydrogen
25 bonding the complementary strand. The displaced homologous or homeologous strand of the recipient helix is guided into the donor helix where it base pairs with its counterpart strand to form a second heteroduplex. The resulting branch point then migrates along the aligned chromosomes thereby elongating and thus stabilising the initial heteroduplexes. Single stranded gaps (if present) are closed by DNA synthesis. The strand cross overs
30 (Holliday junction) are eventually resolved enzymatically to yield the recombination products.

Although in wild type *E. coli* homologous and homeologous recombination are thus mechanistically similar if not identical, homologous recombination in conjugational crosses *E. coli* x *E. coli* occurs five orders of magnitude more frequently than
35 homeologous recombination in conjugational crosses *E. coli* x *S. typhimurium* (Matic et al. 1995; Cell 80, 507-515). The imbalance in favour of homologous recombination was shown to be caused largely by the bacterial MisMatch Repair (MMR) system since its

inactivation increased the frequency of homeologous recombination in *E. coli* up to 1000 fold (Rayssiguier et al. 1989, Nature 342, 396-401).

In *E. coli*, the MMR system (reviewed by Modrich 1991, Annual Rev Genetics 25, 229-253) is composed of only three proteins known as MutS, MutL and MutH. MutS 5 recognizes and binds to base pair mismatches. MutL then forms a stable complex with mismatch bound MutS. This protein complex now activates the MutH intrinsic single stranded endonuclease which nicks the strand containing the misplaced base and thereby prepares the template for DNA repair enzymes.

During recombination, MMR components inhibit homeologous recombination. In 10 vitro experiments demonstrated that MutS in complex with MutL binds to mismatches at the recombination branch point and physically blocks RecA mediated strand exchange and heteroduplex formation (Worth et al., 1994; PNAS 91, 3238-3241). Interestingly, the SOS dependent RuvAB mediated branch migration is insensitive to MutS/MutL, explaining the observed slight increase in SOS dependent homeologous recombination. 15 Homeologous mating even induces the SOS response, thereby taking advantage of RuvAB induction (Matic et al. 1995, Cell 80, 507-515).

The MMR system thus appears to be a genetic guardian over genome stability in *E. coli*. In this role it essentially determines the extent to which genetic isolation, that is, speciation, occurs. The diminished sensitivity of the SOS system to MMR, however, 20 allows (within limits) for rapid genomic changes at times of stress, providing the means for fast adaptation to altered environmental conditions and thus contributing to intraspecies genetic variation and species evolution.

The important role of MMR in preserving genomic integrity has been established also in certain eukaryotes. In its efficiency, the human MMR, for example, may even 25 counteract potential gene therapy tools such as triple-helix forming oligonucleotides including RNA-DNA hybrid molecules (Havre et al., 1993, J. Virology 67: 7234-7331; Wang et al., 1995, Mol. Cell. Biol. 15: 1759-1768; Kotani et al., 1996, Mol. Gen. Genetics 250: 626-634; Cole-Strauss et al., 1996, Science 273: 1387-1389). Such oligonucleotides are designed to introduce single base changes into selected DNA target 30 sequences in order to inactivate for example cancer genes or to restore their normal function. The resulting base mismatches however are recognised by the mismatch repair system which then directs removal of the mismatched base, thereby reducing the efficiency of oligonucleotide induced site-specific mutagenesis.

To date, two families of related genes, homologous to the bacterial *MutS* and *MutL* 35 genes have been identified or isolated in yeast and mammals (recent reviews by Arnheim and Shibata, 1997, Curr. Opinion Genet. Dev. 7, 364-370; Modrich and Lahue, 1996, Annual Rev. Biochem. 65, 101-133; Umar and Kunkel, 1996, Eur. J. Biochem. 238, 297-307). Biochemical and genetic analysis indicated that eukaryotic MutS homologs (MSH)

and MutL homologs (MLH. PMS). respectively. fulfil similar protein functions as their
bacterial counterparts. Their relative abundance. however. could reflect different
mismatch specificity and/or specialisation for different tissues or organelles or
developmental processes such as mitotic versus meiotic recombination.

5   To date, six different genes homologous to MutS have been isolated in yeast
(yMSH). and their homologs have been found in mouse (mMSH) and human (hMSH),
respectively. Encoded proteins yMSH2, yMSH3 and yMSH6 appear to be the main MutS
homologs involved in MMR during mitosis and meiosis in yeast, where the
complementary proteins MSH3 and MSH6 alternatively associate with MSH2 to recognise
10 different mismatch substrates (Masischky et al.. 1996, Genes Dev. 10, 407-420). Similar
protein interactions have been demonstrated for the human homologs hMSH2, hMSH3
and hMSH6 (Acharya et al.. 1996, PNAS 93. 13629-13634).

MutL homologs (MLH and PMS), recently reviewed by Modrich and Lahue (1996,
Annual Rev. Biochem. 65. 101-133) have so far been found in yeast (yMLH1 and
15 yPMS1). mouse (mPMS2) and human (hMLH1. hPMS1 and hPMS2). The hPMS2 is a
member of a family of at least 7 genes (Horii et al.. 1994, Biochem. Biophys. Res.
Commun. 204. 1257-1264) and its gene product is most closely related to yPMS1. Prolla
et al. (1994, Science 265. 1091-1093) presented evidence for yPMS1 and yMLH1 to
physically associate with each other and. together. to interact with the MutS homolog
20 yMSH2 to form a ternary complex involved in mismatch substrate binding.

However, while medical interest in mismatch repair has prompted extensive research
on MMR in bacteria, yeast and mammals, MMR genes have not been isolated from higher
plants prior to the present invention and no attempts to adjust the plant MMR to plant
breeding needs have been reported.

25   SUMMARY OF THE INVENTION

According to a first embodiment of the invention. there is provided an isolated and
purified DNA molecule comprising a polynucleotide sequence encoding a polypeptide
functionally involved in the DNA mismatch repair system of a plant. In one form of this
embodiment, the invention provides an isolated and purified DNA molecule comprising a
30 polynucleotide sequence encoding a polypeptide which is homologous to a mismatch
repair polypeptide of a yeast or of a human. More particularly, the invention provides
polynucleotide sequences encoding polypeptides which are homologous to the mismatch
repair polypeptides MSH3 and MSH6 of Saccharomyces cerevisiae. Still more
particularly, the invention provides the coding sequences of the genes AtMSH3 and
35 AtMSH6 of Arabidopsis thaliana. as defined hereinbelow, and polynucleotide sequences
encoding polypeptides which are homologous to polypeptides encoded by AtMSH3 and
AtMSH6.

According to a second embodiment of the invention, there is provided an isolated and purified polypeptide functionally involved in the DNA mismatch repair system of a plant, for example a polypeptide which is homologous to a mismatch repair polypeptide of a yeast or of a human such as a polypeptide encoded by the genes *AtMSH3* or *AtMSH6* of

5   *Arabidopsis thaliana*, as defined hereinbelow.

According to a third embodiment of the invention, there is provided an isolated and purified DNA molecule comprising a polynucleotide sequence selected from the group consisting of (i) a sequence encoding a polynucleotide which is capable of interfering with the expression of a plant polynucleotide sequence encoding a polypeptide which is

10 homologous to a mismatch repair polypeptide of a yeast or of a human and thereby disabling said plant polynucleotide sequence; and (ii) a sequence encoding a polypeptide capable of disrupting the DNA mismatch repair system of a plant.

According to a fourth embodiment of the invention there is provided a chimeric gene comprising a DNA sequence selected from the group consisting of (i) a sequence encoding

15 a polynucleotide which is capable of interfering with the expression of a plant polynucleotide sequence encoding a polypeptide which is homologous to a mismatch repair polypeptide of a yeast or of a human and thereby disabling said plant polynucleotide sequence, and (ii) a sequence encoding a polypeptide capable of disrupting the DNA mismatch repair system of a plant: together with at least one regulation element capable of

20 functioning in a plant cell. Examples of such regulation elements include constitutive, inducible, tissue type specific and cell type specific promoters such as 35S, NOS, PR1a, AoPR1 and DMC1. Typically, a chimeric gene of the fourth embodiment will also include at least one terminator sequence, more typically exactly one terminator sequence.

In the third and fourth embodiments, said interference, by said polynucleotide

25 sequence, with the expression of a plant polynucleotide sequence encoding a polypeptide which is homologous to a mismatch repair peptide of a yeast or a human typically occurs by hybridisation or by co-suppression.

According to a fifth embodiment of the invention there is provided a plasmid or vector comprising a chimeric gene of the fourth embodiment. A vector of the fifth

30 embodiment may be, for example, a viral vector or a bacterial vector.

According to a sixth embodiment of the invention, there is provided a plant cell stably transformed, transfected or electroporated with a plasmid or vector of the fifth embodiment.

According to seventh embodiment of the invention, there is provided a plant

35 comprising a cell of the sixth embodiment.

According to an eighth embodiment of the invention, there is provided a process for at least partially inactivating a DNA mismatch repair system of a plant cell, comprising

transforming or transfecting said plant cell with a DNA sequence of the third embodiment or a chimeric gene of the fourth embodiment or a plasmid or vector of the fifth embodiment, and causing said DNA sequence to express said polynucleotide or said polypeptide.

5   According to a ninth embodiment of the invention, there is provided a process for increasing genetic variation in a plant comprising obtaining a hybrid plant from a first plant and a second plant, or cells thereof, said first and second plants being genetically different; altering the mismatch repair system in said hybrid plant; permitting said hybrid plant to self-fertilise and produce offspring plants; and screening said offspring plants for 10 plants in which homeologous recombination has occurred.   For example, homeologous recombination may be evidenced by new genetic linkage of a desired characteristic trait or of a gene which contributes to a desired characteristic trait.

According to a tenth embodiment of the invention there is provided a process for obtaining a plant having a desired characteristic, comprising altering the mismatch repair 15 system in a plant, cell or plurality of cells of a plant which does not have said desired characteristic, permitting mutations to persist in said cells to produce mutated plant cells, deriving plants from said mutated plant cells, and screening said plants for a plant having said desired characteristic.

In a preferred form of the ninth and tenth embodiments of the invention, the step of 20 altering the mismatch repair system comprises introducing into said hybrid plant, plant, cell or cells a chimeric gene of the fourth embodiment and permitting the chimeric gene to express a polynucleotide which is capable of interfering with the expression of a plant polynucleotide sequence in a mismatch repair gene of the hybrid plant, plant, cell or cells, or a polypeptide capable of disrupting the DNA mismatch repair system of the hybrid 25 plant or cells.

In other embodiments, the invention provides (a) an oligonucleotide capable of hybridising at 45°C under standard PCR conditions to a DNA molecule of the first embodiment; (b) an oligonucleotide capable of hybridising at 45°C under standard PCR conditions to the DNA of SEQ ID NO: 18 and (c) an oligonucleotide capable of 30 hybridising at 45°C under standard PCR conditions to the DNA of SEQ ID NO:30; with the proviso that the oligonucleotide of (a), (b) and (c) is other than SEQ ID NO:1 or SEQ ID NO:2.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides a diagrammatic representation of the primer sequences used to 35 isolate *AtMSH3*.

Figure 2 is a plasmid map of clone 52, showing restriction enzyme cleavage sites in the 5' half of the full-length cDNA for *AtMSH3*.

Figure 3 is a plasmid map of clone 13, showing restriction enzyme cleavage sites in the 3' half of the full-length cDNA for *AtMSH3*.

Figure 4 is a sequence listing of the coding sequence of *AtMSH3*, together with a deduced sequence of the encoded polypeptide.

5   Figure 5 is a protein alignment of yeast (*Saccharomyces cerevisiae*) and *Arabidopsis thaliana* MSH3 protein.

Figure 6 provides a diagrammatic representation of the primer sequences used to isolate *AtMSH6*.

Figure 7 is a plasmid map of clone 43, showing restriction enzyme cleavage sites in 10 the 5' half of the full-length cDNA for *AtMSH6*.

Figure 8 is a plasmid map of clone 62, showing restriction enzyme cleavage sites in the 3' half of the full-length cDNA for *AtMSH6*.

Figure 9 is a sequence listing of the coding sequence of *AtMSH6*, together with a deduced sequence of the encoded polypeptide.

15   Figure 10 is a protein alignment of yeast (*Saccharomyces cerevisiae*) and *Arabidopsis thaliana* MSH6 protein.

Figure 11 is a genomic sequence listing of *AtMSH6*.

Figure 12 is a plasmid map of plasmid pPF13.

Figure 13 is a plasmid map of plasmid pPF14.

20   Figure 14 is a plasmid map of plasmid pCW186.

Figure 15 is a plasmid map of plasmid pCW187.

Figure 16 is a plasmid map of plasmid pPF66.

Figure 17 is a plasmid map of plasmid pPF57.

Figure 18 is a diagrammatic representation of an antisense gene construction for use 25 in homeologous meiotic recombination.

Figure 19 is a plasmid map of plasmid p3243.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is based on the inventors' discovery that there exist in higher plants genes which are homologous to MMR genes in *E. coli*, and to MMR genes in 30 yeasts and humans.

Thus, the inventors have identified genes, herein designated *AtMSH3* and *AtMSH6*, of the plant *Arabidopsis thaliana* which encode the proteins AtMSH3 and AtMSH6. These plant proteins are homologous to yMSH3 and yMSH6, respectively. The present inventors have isolated cDNAs encoding the proteins AtMSH3 and AtMSH6 and have 35 isolated the complete gene encoding AtMSH6. Given the teaching herein, other genes (for example AtMSH2, and genes of other plants) may be obtained which are involved in DNA mismatch repair in plants, including other genes which encode polypeptides homologous to MMR proteins of yeasts or humans, such as genes which encode

polypeptides homologous to yeast MSH2, MLH1 or PMS2, or to human MLH1, PMS1 or PMS2. For example, given the teaching herein, genes of members of the *Brassicaceae* family or of other unrelated families, for example the *Poaceae*, the *Solanaceae*, the *Asteraceae*, the *Malvaceae*, the *Fabaceae*, the *Linaceae*, the *Canabinaceae*, the *Dauaceae* and the *Cucurbitaceae* family, and which encode polypeptides homologous to MMR proteins of yeasts or humans may be obtained.

Examples of plants whose genes encoding polypeptides homologous to MMR proteins of yeasts or humans may be obtained given the teaching herein include maize, wheat, oats, barley, rice, tomato, potato, tobacco, capsicum, sunflower, lettuce, artichoke, safflower, cotton, okra, beans of many kinds including soybean, peas, melon, squash, cucumber, oilseed rape, broccoli, cauliflower, cabbage, flax, hemp, hops and carrot.

Within the meaning of the present invention, a first polypeptide is defined as homologous to a second polypeptide if the amino acid sequence of the first polypeptide exhibits a similarity of at least 50% on the polypeptide level to the amino acid sequence of the second polypeptide.

A procedure which may be followed to obtain genes *AtMSH3* and *AtMSH6* is described in Example 1. Essentially the same technique may be applied to obtain other mismatch repair genes of *Arabidopsis thaliana*, and essentially the same technique as exemplified herein may be applied to cDNA obtained by reverse transcription of RNA from other plants. Alternatively, given the sequence information disclosed herein, other degenerate oligonucleotide primers, especially oligonucleotides of the invention which are capable of hybridising at 45°C under standard PCR conditions (such as the conditions described in Example 1 using primers UPMU and DOMU) to *AtMSH3* and/or *AtMSH6* may be designed and obtained for use in isolating sequences of plant mismatch repair genes which are homologous to *AtMSH3* or *AtMSH6*, from other plants. Similarly, oligonucleotides of the invention which are capable of hybridising at 45°C under standard PCR conditions to plant mismatch repair genes of plants other than *Arabidopsis thaliana* also fall within the scope of the present invention and may be utilised to obtain mismatch repair genes of still other plants. Typically, such oligonucleotides are capable of hybridising at 45°C under standard PCR conditions to a DNA molecule which encodes a polypeptide which is homologous to a mismatch repair polypeptide of a yeast or a human. The temperature at which oligonucleotides of the invention hybridise to *AtMSH3* and/or *AtMSH6*, or to plant mismatch repair genes of plants other than *Arabidopsis thaliana*, or to DNA molecules which encode polypeptides which are homologous to a mismatch repair polypeptide of a yeast or a human may be higher than 45°C, for example at least 50°C, or at least 55°C, or at least 60°C or as high as 65°C.

The successful gene isolation disclosed herein demonstrates for the first time the existence of MMR in higher plants and indicates the presence of other plant MMR genes. For example, genes encoding the plant homologs of MSH1, MSH2, MSH4, MSH5, PMS1, PMS2 and MLH1 may be identified given the teaching herein. Such genes, as

5 well as those specifically described herein, separately or in combination, are useful in manipulating the plant MMR for plant breeding purposes. Thus, for example, the plant MMR may be altered by including in a plant cell a polynucleotide sequence as defined herein above with reference to the third embodiment of the invention, and causing the polynucleotide sequence to express either a polynucleotide which disables a plant MMR

10 gene, or a polypeptide which disrupts the plant's MMR system.

The DNA molecule of the third embodiment of the invention includes a polynucleotide sequence (herein referred to as a MMR altering gene) which may for example encode sense, antisense or ribozyme molecules characterised by sufficient base sequence similarity or complementarity to the gene to be altered to permit the antisense or

15 ribozyme molecule to hybridise with the plant MMR gene in vivo or to permit the sense molecule to participate in co-suppression. Alternatively, the MMR altering gene may encode a protein or proteins which interfere with the activity of a plant MMR protein and thus disrupt the plant's MMR system. For example, such encoded proteins may be antibodies or other proteins capable of interfering with MMR protein function, such as by

20 complexing with a protein functionally involved in plant MMR thereby disrupting the MMR of the plant. An example of such a protein is the MSH3 protein of *Arabidopsis thaliana* described herein or a protein of another plant which is homologous to the MSH3 protein of *A. thaliana*. For instance, overexpression of MSH3 in a plant cell causes MSH2 present in the cell to be substantially completely complexed, disrupting the

25 mismatch repair mechanism or mechanisms in the cell which are functionally dependent on the presence of a complex of MSH2 with MSH6. Similarly, mismatch repair mechanisms which depend on the presence of a complex of MSH2 and MSH3 may be disrupted by the overexpression of MSH6.

A chimeric gene of the fourth embodiment, incorporating a MMR altering gene,

30 may be prepared by methods which are known in the art. Similarly, the MMR altering gene may be introduced into a plant cell, regenerating tissue or whole plant by techniques known in the art as being suitable for plant transformation, or by crossing. Known transformation techniques include *Agrobacterium tumefaciens* or *A. rhizogenes* mediated gene transfer, ballistic and chemical methods, and electroporation of protoplasts.

35 The MMR altering gene or genes are typically expressed from suitable promoters. Suitable promoters may direct constitutive expression, such as the 35S or the *NOS* promoter. Usually, however, the promoter will direct either inducible or tissue specific (e.g. callus; embryonic tissue; etc.), cell type specific (e.g. protoplasts; meiocytes; etc.) or developmental (e.g. embryo) expression of the altering gene or genes, in order for the

MMR system to function in tissue types or cell types, or at developmental stages of the plant, in which it is not desirable for the MMR system to be altered. Using such promoters, therefore, the activity of a MMR altering gene may be limited to a specific stage during plant development or it may be altered by controlling conditions external to

5 the plant, and the deleterious effects of a permanently disabled or altered DNA mismatch repair system in a plant may be avoided. Examples of suitable promoters which are not constitutive are known in the art and include inducible promoters such as *PR*1a (reviewed by Gatz, 1997, Annual Rev. Plant Phys. Plant Mol. Biol. 48: 89), tissue specific promoters such as *AoPR1* (Sabahattin et al., 1993, Biotechnology 11: 218), and cell-type

10 specific promoters such as *DMC1*.

The process of the tenth embodiment of the invention provides, for example, a chimeric gene in accordance with the invention may further be physically linked to one or more selection markers such as genes which confer phenotypic traits such as herbicide resistance, antibiotic resistance or disease resistance, or which confer some other recognisable trait such as male sterility, male fertility, grain size, colour, growth

15 rate, flowering time, ripening time, etc.

The process of the tenth embodiment of the invention provides, for example, a process for generating intraspecies genetic variation by altering the mismatch repair system in a plant cell, in regenerating plant tissue or in a whole plant. The plant cell, regenerating tissue or whole plant includes and expresses one or more MMR altering

20 genes which are capable of altering mismatch repair in the plant cell, regenerating tissue or whole plant. Alteration of MMR may be achieved, for example, by inactivating the genes encoding plant MSH3 and/or plant MSH6. It is preferred to inactivate the plant MSH3 and MSH6 encoding genes at the same time and in the same plant cell, regenerating tissue or whole plant. Typically in this preferred form of the invention

25 inactivation of either plant MSH3 or MSH6 alone is insufficient to substantially alter the plant's mismatch repair system and only when both MSH3 and MSH6 are inactivated simultaneously is the plant's mismatch repair system sufficiently altered to prevent the MMR system from recognising base pair mismatches, base insertions or deletions as a result of DNA replication errors, DNA damage, or oligonucleotide induced site-specific

30 mutagenesis. However, in some applications of the invention, inactivation of only one gene may also be used to cause genomic instability or increase the efficiency of site-specific mutagenesis.

If desired, the MMR altering gene or genes may later be rendered non-functional or ineffective, or may be removed from the genome of the plant cell, regenerating tissue or

35 whole plant in order to restore mismatch repair in the plant cell, regenerating tissue or whole plant. The MMR altering gene or genes may be inactivated by means of known gene inactivation tools, such as ribozymes, or may be removed from the genome using gene elimination systems known in the art, such as *CRE/LOX*. It is preferred to render two genes, whose gene products combine to incapacitate MMR, ineffective by separating

the altering genes through segregation. Therefore, in a preferred embodiment of the invention a first plant cell or plant is generated in which only plant *MSH3* is incapacitated, and a second plant cell or plant is generated in which only plant *MSH6* is incapacitated. The combination of both genomes, for example by crossing, then produces significant

5 MMR deficiency in those cells or plants which have inherited both altering genes. If the altering genes are expressed from unlinked loci, gene segregation restores MMR activity in the progeny of the cells or plants.

In a process of the ninth embodiment of this invention, homeologous recombination is enhanced between different genomes, chromosomes or genes in plant cells or plants by

10 altering MMR in said plant cells or plants. Such genomes, chromosomes or genes are characterised in that they originate from different plant families, genera, species, subspecies, plant varieties or lines. Hybrid plant cells or hybrid plants may be produced by crossing, by cell fusion or by other techniques known in the art. These plant cells or plants are further characterised by expressing one or more genes that are capable of

15 altering mismatch repair in the plant cell or plants.

In the process of the ninth embodiment, the homeologous recombination is typically for the purpose of introducing a desired characteristic in the hybrid plant. In this typical application of the process of the ninth embodiment, and in the process of the tenth embodiment the desired characteristic may be any characteristic which is of value to the

20 plant breeder. Examples of such characteristics are well known in the art and include altered composition or quality of leaf or seed derived storage products (e.g. oil, starch, protein), altered composition or quality of cell walls (e.g. decrease in lignin content), altered growth rate, prolonged flowering, increased plant yield or grain yield, altered plant morphology, resistence to pathogens, tolerance to or improved performance under

25 environmental stresses of various kinds, etc.

In a preferred form of the tenth embodiment, an MMR altering gene is co-introduced along with the homeologous genome, chromosome or gene of another plant cell or plant into an MMR proficient plant cell or MMR proficient plant to produce a hybrid plant cell or hybrid plant in which homeologous recombination can occur.

30 Suitably, the MMR proficient plant cell or MMR proficient plant may also include an MMR altering gene. For example a gene capable of inactivating plant MSH3 may be co-introduced along with the homeologous genome, chromosome or gene of another plant cell or plant into an MMR proficient plant cell or MMR proficient plant in which MSH6 is inactivated. A resultant hybrid plant in which homeologous recombination occurs will

35 include both the MSH3 and MSH6 altering genes and its MMR system will therefore be inactivated.

In this form of the invention, if hybrid plants are to be produced by crossing, the MMR altering gene preferably originates from the male parent, thus ensuring that the

MMR altering gene is always introduced and is not present in the recipient cell. That is, the MMR of the recipient cell, prior to introduction of the MMR altering gene, is typically proficient. Alternatively, if an MMR altering gene is present in a recipient cell it may be ineffective or inefficient on its own, or it may be linked to an inducible or tissue specific or cell type specific promoter which only renders the MMR altering gene active under limited conditions.

Thus, in a preferred form of the process of the ninth embodiment, the MMR system of the hybrid plant is initially unaltered. In this form of the process, the step of altering the mismatch repair system may comprise introducing into the hybrid plant, or cells thereof, a MMR altering gene, such as by *Agrobacterium tumefaciens* or *A. rhizogenes* mediated gene transfer, ballistic and chemical methods, and electroporation of protoplasts.

The MMR altering gene or genes are typically expressed from suitable promoters, as described above. Preferably, the promoter is transcriptionally active in mitotically and meiotically active tissue and/or cells to ensure MMR alteration after chromosome pairing at mitosis and meiosis, respectively. The preferred timing for MMR alteration is at meiosis, because recombinant genomes, chromosomes or genes are directly transmitted to the progeny. A suitable meiocyte specific promoter is for example the *DMC1* promoter from *Arabidopsis thaliana* ssp. *Ler.* (Klimyuk and Jones, 1997, Plant J. 11, 1-14). However, mitotic homeologous recombination is also a desirable outcome as somatic recombination events can be transmitted to offspring due to the totipotency of plant cells and the lack of predetermined germ cells in plants.

If desired, the MMR altering gene or genes may later be rendered non-functional or ineffective, or may be removed from the hybrid plant or hybrid plant cells, in order to restore mismatch repair in the hybrid plant or hybrid plant cells. The MMR altering gene or genes may be inactivated by means of known gene inactivation tools as described herein above.

EXAMPLES

Example 1.   Cloning of the *AtMSH3* and *AtMSH6* coding sequences

Isolation of partial *AtMSH3* and *AtMSH6* consensus sequences

Degenerate oligonucleotides UPMU (SEQ ID NO:1) and DOMU (SEQ ID NO:2)

UPMU   CTGGATCCACIGGICCIAA(C/T)ATG
DOMU   CTGGATCC(A/G)TA(A/G)TGIGTI(A/G)C(A/G)AA

were used to isolate *AtMSH3* and *AtMSH6* sequences by PCR amplification.

Primers UPMU and DOMU correspond to conserved amino acid sequences of the proteins MutS (*E. coli* and *S. typhimurium*), HexA (*S. pneumoniae*), Rep1 (mouse) and Duc1 (human). The conserved regions to which they are targeted are TGPNM for UPMU (amino acid positions 852-856 for AtMSH6 and 816-820 for AtMSH3) FATHY or FVTHY

13

for DOMU (amino acid positions 964-968 for AtMSH6 and 928-932 for AtMSH3, respectively.) These primers have been used to isolate MSH2 and MSH1 from yeast (Reenan and Kolodner, Genetics 132: 963-973 (1992)) and MSH2 from *Xenopus* and mouse (Varlet et al., Nuc. Acids Res. 22:5723-5728 (1994)).

5   Template single strand cDNA was produced by reverse transcription of 2 µg total RNA from a cell suspension culture of *Arabidopsis thaliana* ecotype Columbia (Axelos et al. 1989, Mol. Gen. Genetics 219: 106-112). The PCR reaction was performed under the following conditions in a final volume of 100µl: 0.2mM dNTP, 1µM each primer, 1XPCR buffer, 1u *Taq* DNA polymerase (Appligene) in the presence of template cDNA. PCR
10   parameters were 5 minutes at 94°C, followed by 30 cycles of 40 seconds at 95°C, 90 seconds at 45°C, 1 minute at 72°C. The amplification products were cloned into pGEM-T vector (Promega) and sequenced. Two different clones were isolated, S5 (350bp) was homologous to *MSH3*, S8 (327bp) was homologous to *MSH6*. Complete cDNA sequences were then isolated according to the Marathon cDNA amplification kit procedure (Clontech).
15   In summary, this procedure involves producing double stranded cDNA by reverse transcription of 2µg polyA+ RNA from the cell suspension culture of *Arabidopsis*. Adaptors are ligated on each side of the cDNA. The ligated cDNA is used as a template for 5' and 3' RACE PCR reactions in the presence of primers that are specific for the adaptor on one side (AP1 and AP2), and specific for the targeted gene on the other side. A 5' and a 3'
20   fragment that overlap are thus produced for each gene. The complete gene coding sequence can be reconstituted taking advantage of a unique restriction site, if available, in the overlapping region. Specific details of this procedure as it was used to isolate *AtMSH3* and *AtMSH6* coding regions, are as follows.

Isolation of *AtMSH3* complete coding sequence

25   From the sequence of clone S5, primer 636 (SEQ ID NO:3) was designed:
      636   TGCTAGTGCCTCTTGCAAGCTCAT.

Primer AP1 (SEQ ID NO:4) is complementary to a portion of an adaptor sequence which had been ligated to the 5' and 3' ends of *Arabidopsis* cDNA:
      AP1   CCATCCTAATACGACTCACTATAGGGC.

30   PCR performed on the ligated cDNA with primers 636 and AP1 for the 5' RACE PCR was followed by a second round of amplification with the nested primers AP2 (SEQ ID NO:5) and S525 (SEQ ID NO:6)
      AP2   ACTCACTATAGGGCTCGAGCGGC
      S525   AGGTTCTGATTATGTGTGACGCTTTACTTA

35   (the latter was also designed to correspond to a part of the sequence of clone S5) and produced a 2720bp DNA fragment. Figure 1 provides a diagrammatic representation of the primer sequences used to isolate *AtMSH3*. Another primer (S51, SEQ ID NO:7)
      S51   GGATCGGGTACTGGGTTTTGAGTGTGAGG

was designed closer to the 5' border and permitted the determination of 99bp upstream to the ATG initiation codon. For the 3' RACE PCR, a first PCR reaction was performed with primers AP1 and 635 (SEQ ID NO:8).

    635   GCACGTGCTTGATGGTGTTTTCAC

5 followed by a second round of amplification, using the nested primers AP2 and S523 (SEQ ID NO:9)

    S523   TCAGACAGTATCCAGCATGGCAGAAGTA

which produced a DNA fragment of 890bp. Both DNA fragments were subcloned into pGEM-T and sequenced. Since PCR amplification using the Expand Long Template PCR

10 System (Boehringer-Mannheim) produced errors in the sequence, new oligonucleotides were designed to isolate those sequences again by PCR, but with the high fidelity DNA polymerase *Pfu*. PCR with primers 1S5 (SEQ ID NO:10) and S53 (SEQ ID NO:11)

    1S5   ATCCCGGGATGGGCAAGCAAAAGCAGCAGACGA

    S53   GACAAAGAGCGAAATGAGGCCCCTTGG

15 amplified the 1244bp fragment clone 52 (SEQ ID NO:12, cloned into pUC18/Sma1). PCR with primers S52 (SEQ ID NO:13) and 2S5 (SEQ ID NO:14)

    2S5   ATCCCGGGTCAAAATGAACAAGTTGGTTTTAGTC

    S52   GCCACATCTGACTGTTCAAGCCCTCGC

amplified the 2104bp clone 13 (SEQ ID NO:15, cloned into pUC18/Sma1). The complete

20 coding sequence of the *AtMSH3* gene was reconstructed in pUC18 by ligating the 5' half of *AtMSH3* (clone 52) to the 3' half of *AtMSH3* (clone 13) after digesting with *Bam*H1 which has a unique cleavage site in the overlapping region of both clones. This manipulation yielded plasmid pPF26. The *Sma*I fragment from pPF26 contains the complete *AtMSH3* coding sequence. The remaining primers referred to in Figure 1 are as

25 follows:

    S51   GGATCGGGTACTGGGTTTTGAGTGTGAGG (SEQ ID NO:16)

    S525   AGGTTCTGATTATGTGTGACGCTTTACTTA (SEQ ID NO:17)

       Figures 2 and 3 provide plasmid maps of clones 52 and 13 respectively, showing restriction enzyme cleavage sites. The complete *AtMSH3* coding sequence (SEQ ID NO:18)

30 is 3246bp long and is shown in Figure 4 together with the deduced sequence (SEQ ID NO:19) of the encoded polypeptide. *AtMSH3* is clearly homologous to the yeast and mouse *MSH3* genes. A sequence alignment of polypeptides encoded by *AtMSH3* and that encoded by *Saccharomyces cerevisiae MSH3* is set out in Figure 5.

Isolation of the *AtMSH6* complete coding sequence and genomic sequences

35   The same procedure allowed isolation of the *AtMSH6* cDNA. Figure 6 provides a diagrammatic representation of the primer sequences used to isolate *AtMSH6*. For the 5' RACE PCR, primers 638 (SEQ ID NO:20) and AP1 (SEQ ID NO:4)

    638   TCTCTACCAGGTGACGAAAAACCG

allowed the amplification of a 2889 DNA fragment. Primer S81 (SEQ ID NO:21)

S81   CGTCGCCTTTAGCATCCCCTTCCTTCAC

helped define the 142bp upstream to the ATG initiation codon. On the 3' side. RACE PCR
was initially performed with primers S823 (SEQ ID NO:22) and AP1 (SEQ ID NO:4).

S823   GCTTGGCGCATCTAATAGAATCATGACAGG

and then with the nested primers 637 (SEQ ID NO:23) and AP2 (SEQ ID NO:5).

637   GACAGCGTCAGTTCTTCAGAATGC

to produce a 774bp DNA fragment. As for *AtMSH3*, those fragments were cloned and
sequenced. Re-isolation of the DNA sequence using the high fidelity *Pfu* polymerase and
newly designed primers 1S8 (SEQ ID NO:24) and S83 (SEQ ID NO:25) (for the 5' side) led
to a 2182 bp DNA fragment identified as clone 43 (SEQ ID NO:26, cloned in
pUC18/SmaI), and a 1379bp clone identified as clone 62 (SEQ ID NO:27, also cloned in
pUC18/SmaI).

1S8   ATCCCGGGATGCAGCGCCAGAGATCGATTTTGT

2S8   ATCCCGGGTTATTTGGGAACACAGTAAGAGGATT (SEQ ID
NO:28)

S82   GCGTTCGATCATCAGCCTCTGTGTTGC (SEQ ID NO:29)

S83   CGCTATCTATGGCTGCTTCGAATTGAG

Figures 7 and 8 provide plasmid maps of clones 43 and 62 respectively. showing restriction
enzyme cleavage sites. Clones 43 and 62 were digested by the *Xmn*1 restriction enzyme for
which a unique site is present in their overlapping region and then ligated. The complete
AtMSH6 coding sequence (SEQ ID NO:30) is 3330bp long and is shown in Figure 9
together with the deduced sequence (SEQ ID NO:31) of the encoded polypeptide. *AtMSH6*
is clearly homologous to the yeast and mouse *MSH6*genes. A sequence alignment of
polypeptides encoded by *AtMSH6* and that encoded by *Saccharomyces cerevisiae MSH6* is
set out in Figure 10.

An *AtMSH6* genomic sequence was also isolated from a genomic DNA library
constituted after partial *Sau3*AI digestion of DNA from the *Arabidopsis* cell suspension.
8062bp were sequenced that covered the *AtMSH6* gene and show colinearity with the
cDNA. 16 introns are found scattered along the gene. The complete genomic sequence
(SEQ ID NO:98) is shown in Figure 11.

Example 2.   A measure of somatic variation in MMR deficient plants

Constructs

        Constructs with antisense *AtMSH3* or antisense *AtMSH6* or both *AtMSH3/AtMSH6*
under the control of a single 35S promoter have been inserted into the binary vector
pPZP121 (Hajdukiewicz et al., Plant Mol. Biol. 23. 793-799) between the right and left
borders of the T-DNA. The pPZP121 plasmid confers chloramphenicol resistance to
Escherichia coli or *Agrobacterium tumefaciens* bacteria. The *aacC1* gene is carried by the
T-DNA and allows selection of transformed plant cells on gentamycin (Hajdukiewicz et al.,
Plant Mol. Biol. 25, 989-994). For the purpose of expressing antisense constructs. a 35S

promoter/terminator cassette with a polylinker was introduced into pPZP121. The 3' ends of the considered genes have been chosen since this region seems more efficient for antisense inhibition. For *AtMSH3* this corresponds to clone 13 (2104bp), for *AtMSH6* this corresponds to clone 62 (1379bp). Clone 13 comprises 2104bp of the 3' region that were cut off the pUC18 vector by SaI1/Sst1 restriction, blunted with T4 DNA polymerase and ligated into the T4 DNA polymerase blunted *Bam*HI site of pPZP121/35S, creating clone pPF13. The same procedure was followed for the 3' region of *AtMSH6* clone 62 (1379bp) thus creating plasmid pPF14. For the double constructs, the 3' region (from clone 62) of *AtMSH6* was introduced ahead of the *AtMSH3* region into pPF13 creating pCW186 and reciprocally, the 3' region of *AtMSH3* (from clone 13) was introduced ahead of *AtMSH6* into pPF14, creating pCW187.

These constructs were introduced into the Arabidopsis cells (as described below) of wildtype Columbia and of the Columbia tester line.

An alternative strategy to antisense inhibition of *AtMSH6* comes from experiments of Marra et al. (1998. Proc. Natl. Acad. Sci USA 95. *8568-8573)* who show that overexpression of functional *MSH3* results in depletion of MSH6 protein in human cells. This depletion may generate a mismatch repair mutant phenotype.

For the purpose of overexpressing functional *AtMSH3* protein in plant cells, the complete *MSH3* coding region was excised from pPF26 (example 1) by digestion with *Sma*I, and was inserted into the *Sma*I site of pCW164. The resulting construct was named pPF66. It contains a complete *AtMSH3* gene under the control of the *35S* promoter inside the left (LB) and right (RB) border of the T-DNA. This T-DNA also contains the *hpt2* gene for gentamycin selection. Plasmid pPF66 was introduced into Arabidopsis cells as described below. One cell clone was selected which clearly overexpressed the *AtMSH3* gene as shown by Northern analysis. Figures 12-16 provide plasmid maps of plasmids pPF13, pPF14, pCW186, pCW187 and pPF66, respectively.

Construction of tester construct

For the purpose of Forward Mutagenesis Assays, a tester construct was built containing the coding regions for *npt*II, *cod*A, *uid*A. All three genes are driven by the *35S* promoter and are terminated by the *35S* terminator. This construct was obtained by introducing an EcoR1 fragment encoding the *cod*A cassette (2.5kb) and a *Hind*III fragment encoding the *uid*A *(GUS)* cassette (2.4kb) into the pPZP111 vector (Hajdukiewicz et al.,1994, Plant Mol Biol 23: 793-799) which already contained the *npt*II expression cassette. This new plasmid was named pPF57. *Npt*II is used to select for transformed plant cells. *GUS* is used to analyse the degree of gene silencing in the construct (i.e. to identify cell lines in which the transgenes are expressed), and *cod*A is used as a marker for forward mutagenesis (described below).

The plasmid map of pPF57 is provided in Figure 17.

Plant cell transformation

The constructs are introduced into *Agrobacterium* by electroporation. Plant cells are then transformed by co-cultivation. A suspension culture of *Arabidopsis thaliana* cells that has been established by Axelos et al. (1992, Plant Physiol. Biochem. 30, 1-6) may be used. One day old freshly subcultured cells are diluted five times into AT medium (Gamborg B5 medium, 30g/l sucrose, 200µg/l NAA). 10µl of saturated *Agrobacterium* containing the transforming T-DNA constructs are added to 10ml diluted cells in a 100ml erlenmeyer. The co-cultivation is agitated slowly (80rpm) for 2 days. The cells are then washed 3 times into AT medium and finally resuspended in the same initial volume (10ml). The culture is agitated for 3 days to allow expression before plating on selection plates (AT/BactoAgar 8g/l+gentamycin 50µg/ml). Transformed individual calli are isolated 3 weeks later.

Tester Strain

The tester construct on plasmid pPF57 was introduced into Arabidopsis cells of wildtype Columbia using the transformation protocol described above. Among 10 candidate transformants, one cell clone was shown (by Southern analysis) to have a unique T-DNA insertion. All three genes were shown to be functional in this cell line as indicated by resistance to kanamycin, blue staining in the presence of X-Glu *(GUS)*, and sensitivity to 5-fluoro-cytosine *(codA)*.

MMR altering genes (described above) were then introduced individually into the tester line and transformed cells are used for analysis of both Microsatellite Instability and Forward Mutagenesis.

Microsatellite analysis

Microsatellites have been described in *Arabidopsis* (Bell and Ecker, 1994, Genomics 19. 137-144). The present Example is based on a study of instability of microsatellites that are polymorphic amongst different ecotypes. DNA is extracted from the transformed calli. Specific primers have been defined that are used to amplify the microsatellite sequence. One of the two primers is previously P^{32} labelled by T4 kinase. In case of a polymorphic variation, new PCR products appear that do not follow the expected pattern of migration on a polyacrylamide gel. This is a commonly observed feature for MMR deficient cells in yeast or mammalian cells.

In particular, the present Example describes a study on microsatellites ca72 (CA_{18}), nga172 (GA_{29}). and ATHGENEA(A_{39}), chosen because they belong to the types predominantly affected in human mismatch repair deficient tumors. The size of these microsatellites is not conserved from one *Arabidopsis* ecotype to the other.

Arabidopsis cells which are transformed with an MMR altering gene (above) and control cells not expressing the MMR altering gene are allowed to form calli. DNA is

rapidly extracted from the calli and is analysed for microsatellite instability as described in detail by Bell and Ecker 1994, Genomics 19, 137-144. In summary, the relevant microsatellite is amplified by PCR using P32 labelled primers. The PCR products are separated on a DNA sequencing gel for size determination. Size differences between microsatellites from transformed and control cells not expressing the MMR altering gene in question indicate microsatellite instability as a result of MMR alteration.

The sequences of primers used for PCR amplification of microsatellites ca72 and nga172 are included in Table 1. PCR amplification of microsatellite ATHGENEA made use of a forward primer containing the sequence

ACCATGCATAGCTTAAACTTCTTG (SEQ ID NO:32)

and of a reverse primer containing the sequence

ACATAACCACAAATAGGGGTGC (SEQ ID NO:33).

The amplification for microsatellite ca72 revealed in Columbia control cells (with respect to the MMR altering gene) a 248 bp long PCR fragment instead of the published length of 124 bp. DNA sequencing verified this fragment as a CA_{18} microsatellite.

Forward mutagenesis assay

Tester cells transformed with antisense *AtMSH3* or antisense *AtMSH6* or both *AtMSH3/AtMSH6* are analysed for the stability of the *codA* gene. The functional *codA* gene confers to sensitivity to 5-fluoro-cytosine (5FC), whereas a gene inactivating mutation in *codA* will confer resistance to 5FC. The frequency of resistant cells is therefore a good indicator of somatic variation as a direct result of MMR alteration. Variants resistant to 5FC are first analysed for GUS activity. If GUS is inactive, 5FC resistance is assumed to be due to gene silencing (all three genes are under the *35S* promoter). If GUS is active, 5FC resistance is assumed to be due to forward mutations that have inactivated *codA*. PCR is then performed on the putative *codA* mutant genes which is then sequenced to confirm the presence of forward mutations in *codA*.

Besides *codA*, other marker genes may also be used for the Forward Mutagenesis Assay such as the *ALS* gene (conferring sensitivity to valine or to sulfonylurea; Hervieu and Vaucheret, 1996, Mol. Gen. Genet. 251 220-224; Mazur et al. 1987, Plant Physiol. 85 1110-1117).

Example 3.   Homeologous meiotic recombination in *Arabidopsis thaliana*

A.   Construction of a meiocyte specific gene expression cassette comprising the *DMC1* promoter and the *NOS* terminator

(i) The *DMC1* promoter may be used as published by Klimyuk and Jones, 1997, Plant J. 11.1-14). To obtain a more convenient alternative for gene cloning, a 3.3 Kb

long subfragment of the *DMC1* promoter was obtained by PCR from genomic DNA of *Arabidopsis thaliana* (ssp. Landsberg erecta "*Ler*").

The PCR was done in three rounds:

Round One: A 3.7 Kb long product was obtained using the forward primer DMCIN-A comprising the sequence

GAAGCGATATTGTTCGTG (SEQ ID NO:34)

and the reverse primer DMCIN-B comprising the sequence

AGATTGCGAGAACATTCC (SEQ ID NO:35).

The weak amplification product was then used as template for round two and three.

Round Two: A 3.1 Kb long product comprising the promoter and the 5' untranslated leader was obtained using forward primer DMCIN-1, which contained the sequence

acgcgtcgacTCAGCTATGAGATTACTCGTG (SEQ ID NO:36)

and introduced a *SalI* cloning site at the 5' end of the promoter fragment, and reverse primer DMCIN-2 which contained the sequence

gctctagaTTTCTCGCTCTAAGACTCTCT (SEQ ID NO:37)

and introduced a *XbaI* site at the 3' end of the PCR fragment.

Round Three: A 0.2 Kb long product comprising the first exon/intron of the *DMC1* promoter was obtained using forward primer DMCIN-3, which contained the sequence

gctctagaGCTTCTCTTAAGTAAGTGATTGAT (SEQ ID NO:38)

and introduced a *XbaI* site at the 5' end of the PCR fragment, and reverse primer DMCIN-4, containing the sequence

tcccccgggctcgagagatctccatggTTTCTTCAGCTCTATGAATCC (SEQ ID NO:39)

and introduced at the 3' end of the PCR product restriction sites for *NcoI*, *BglII*, *XhoI* and *SmaI*.

The products obtained in round Two and Three were digested with *XbaI* and subsequently ligated to reconstitute a 3.3 Kb long *DMC1* promoter from which the first two in-frame ATG start codons were replaced with a unique restriction site for *XbaI*. This promoter can be cloned between the restriction sites for *SalI* and *SmaI* of p3264, which contains the *SacI-EcoRI* NOS terminator in pBIN19, to yield the entire expression cassette in pBIN19. This cassette contains the following cloning sites: *NcoI*, *BglII*, *XhoI*, *SmaI* and (already present on p3264) *KpnI* and *SacI*.

(ii) Another strategy yielded the following convenient *DMC1* promoter. A 1.8 kb DNA fragment comprising the 3' terminal part of the meiocyte specific *DMC1* promoter was isolated by PCR from purified genomic DNA of *Arabidopsis thaliana* (ssp. Landsberg erecta "*Ler*"). The forward PCR primer (DMC1a) contained the sequence

acgcgtcgacGAATTCGCAAGTGGGG (SEQ ID NO:40)

and introduced a *SalI* cloning site at the 5' end of the promoter fragment. The reverse PCR primer (DMC1b) contained the sequence

tccatggagatctcccgggtacCGATTTGCTTCGAGGG (SEQ ID NO:41)

introducing a polylinker region at the 3' end of the promoter fragment. The PCR reaction was carried out with VENT DNA Polymerase (NEB) over 25 cycles using the following cycling protocol: 1 minute at 94°C, 1 minute at 54°C, 2 minutes at 72°C.

5   The PCR reaction yielded a blunt ended DNA fragment which was digested with restriction enzyme SalI and was cloned into the cleavage sites of restriction enzymes SalI and SmaI in plasmid p2030, a pUC118 derivative containing the SacI-EcoRI NOS terminator fragment of pBIN121. The cloning yielded plasmid p2031, containing the DMC1 promoter-polylinker-NOS terminator expression cassette depicted in Figure 18.

10 B.   Construction of an MSH3 antisense gene under the control of the DMC1 promoter

A 2.1 kb DNA fragment encoding the carboxyterminal part of AtMSH3 was removed from the polylinker of clone 13 described in Example 1 after (i) digestion with KpnI, (ii) blunting of the DNA ends generated by KpnI and (iii) digestion with BamHI. The isolated fragment was then cloned in antisense orientation downstream of the DMC1
15 promoter in plasmid p2031, which had been digested with restriction enzymes SmaI and BglII. This cloning yielded plasmid p2033 (Figure 18).

After digestion of p2033 with EcoRI, a 4.1 kb DNA fragment was recovered comprising the DMC1 promoter, the partial MSH3 cDNA in antisense orientation with respect to the promoter and the NOS terminator. This fragment was cloned into the EcoRI
20 restriction site of plant transformation vector pNOS-Hyg-SCV to yield plasmid p3242 (Figure 18).

C.   Construction of a combined MSH6/MSH3 antisense gene under the control of a single DMC1 promoter

A 3.1 kb fragment, encoding in antisense orientation the partial AtMSH6 and AtMSH3
25 sequences and the 35S terminator, was isolated from pCW186 by digestion with KpnI. The fragment was treated with Klenow enzyme to blunt both ends. It was then cloned into the SmaI site of plasmid p3243 (a pNOS-Hyg-SCV derivative, illustrated in Figure 19), which had been opened to delete the region between the SmaI sites. Clones containing the fragment in the antisense orientation with respect to the DMC1 promoter (described in
30 A(ii) above) were identified by diagnostic digestion with BamHI. The selected construct was named p3261.

Another practical way of cloning the double antisense gene is as follows. A 1 kb DNA fragment encoding the carboxyterminal part of AtMSH6 is isolated from clone 62 described in Example 1 after digestion of clone 62 plasmid DNA with BamHI, which
35 cleaves in the 5' polylinker region flanking the partial cDNA, and with EcoRI, which cleaves within the cDNA. The isolated fragment is treated with Klenow enzyme to blunt both its ends and is cloned into the recipient plasmid p2033 or p3242. For the purpose of

cloning. the recipient plasmid may be cleaved with either *Ava*I or *Nco*I and can be blunted with *Klenow* enzyme to produce blunt acceptor ends for fragment cloning. This cloning yields two opposite orientations of cloned fragment DNA with respect to the *DMC*1 promoter. These can be identified by diagnostic digestion with restriction enzymes *Sca*I or *Xmn*I in conjunction with *Sac*I. The selected construct contains the *DMC*1 promoter, the combined partial cDNAs for *AtMSH3* and *AtMSH6* (both cloned in antisense orientation with respect to the *DMC*1 promoter) and the *NOS* terminator. If the recipient plasmid is p2033, the combined antisense gene under control the single *DMC*1 promoter is recovered from the vector after *Eco*RI digestion and is cloned into the *Eco*RI restriction site of pNOS-Hyg-SCV.

D. Construction of a full-length *MSH3* sense gene under control of the *DMC*1 promoter for overexpression of functional MSH3 protein

Overexpression of MSH3 protein was shown in human cells (Marra et al., 1998, Proc. Natl. Acad. Sci. USA 95, 8568-8573) to complex all available *MSH2* protein. This leaves *MSH6* protein without its partner, leading to the degradation of *MSH6* protein and eventually to a mismatch repair phenotype.

This phenomenon is exploited to increase homeologous meiotic recombination in Arabidopsis as an alternative to antisense inhibition of *AtMSH6*. For this purpose the full-length cDNA encoding *AtMSH3* is isolated from plasmid pPF66 by digestion with *Sma*I and is cloned into the *Sma*I site of the *DMC*1 expression cassettes described in A(i).

E. Selection of Recombination markers on homeologous chromosomes of *Arabidopsis thaliana* subspecies *Landsberg erecta* (Ler), *Columbia* (Col) and *C24*, respectively

E(i). Visual recombination markers in *Arabidopsis th.* subspecies *C24*:

Agrobacterium mediated transformation with a T-DNA containing a *35S-GUS* gene, inactivated by insertion of a *35S-Ac* transposable element (Finnegan et al., 1993, Plant Mol. Biol. 22, 625-633), had yielded a *C24* line in which the T-DNA construct was integrated into chromosome 2. Genetic and molecular analysis of this line shows that the *Ac* transposon had excised from its T-DNA locus thereby restoring *GUS* activity, but had re-inserted into the chromosome at a distance of 16.4 cM, where it stayed fixed (due to disablement of *Ac*) within the *chlorina* gene. Insertional inactivation of the *chlorina* gene caused a bleached phenotype in those plants that were homozygous for this mutation. Because of the two linked phenotypic markers, *chlorina3A:Ac* and *GUS*, this *C24* line was used in crosses to wildtype Ler for analysis of meiotic homeologous recombination on chromosome 2 in conjunction with molecular recombination markers.

E(ii). Visual recombination markers in *Arabidopsis th. Ler*:

The Ler line NW1 (obtained from NASC, Nottingham, UK) contains one recessive visual marker per chromosome. i.e. *an-1* on Chr.1, *py-1* on Chr.2, *gl1-1* on Chr.3, *cer2-1*

on Chr.4, and *ms1-1* on Chr.5. This line is used in crosses to wildtype *C24* which expresses an MMR altering gene for analysis of meiotic homeologous recombination on chromosomes 1-5 in conjunction with molecular recombination markers listed in Table 1.

Other *Ler* lines from NASC have several visual markers in close proximity to each other on the same chromosome. When these lines are used for hybrid production, analysis of homeologous meiotic recombination may make use entirely of visual recombination markers. Particularly suitable for crossing to *C24* wildtype that is expressing a MMR altering gene are the following *Ler* lines:

NW22: relative markers are *dis1* - (4 cM) - *ga4* - (11 cM) - *th1* on chromosome 1.

NW10: relevant markers are *tz-201* - (5 cM) - *cer3* on chromosome 5.

NW134, relevant markers are *ttg* - (4 cM) - *ga3* on chromosome 5.

NW24 (*abi3*-1) and NW64 (*gl1*-1). When present in the same plant on chromosome 3, *abi3*-1 and *gl1*-1 are 13 cM apart. Since this marker combination is not available from NASC, we have combined these markers by crossing line NW24 to line NW64. The F1 offspring were allowed to self-fertilise and to produce F2 seeds. F2 Plants which carry both markers as homozygous traits on the same chromosome can be identified firstly, by germinating F2 seeds on germination medium containing selective concentrations of abscisic acid, and subsequently, by identifying among the abscisic acid resistant plants those individuals which show the glabra phenotype.

E(iii)   Molecular recombination markers in *Col, Ler* and *C24*:

The genome of *Arabidopsis thaliana* is interspersed with unique base sequences arranged as simple tandem repeats. Allelic repeats can vary in length between different *Arabidopsis* subspecies and when amplified by PCR yield diagnostic DNA products of different length named Simple Sequence Length Polymorphisms (SSLPs). Many SSLPs have been genetically mapped and have been assigned to unique chromosome locations on the recombinant inbred map (Bell and Ecker, 1994, Genomics 19, 137-144; Lister and Deans lines, Weeds World 4i, May 1997).

In Table 1 are listed 28 mapped and established SSLPs between *Ler* and *Col*. A number of PCR primer pairs are described herein (SEQ ID NO:42 to SEQ ID NO:97) which also yielded SSLPs between *C24* and *Ler* (19 SSLPs) or between *C24* and *Col* (25 SSLPs), respectively. Polymorphic SSLPs can be used as molecular markers in the analysis of homeologous recombination between genomes from these subspecies.

The PCR reactions (25 μL) were carried out over 35 cycles (15 seconds at 94°C, 30 seconds at 55°C and 30 seconds at 72°C), with 0.25 U Taq DNA polymerase and 0.6 μg genomic DNA in reaction buffer containing 2 mM $MgCl_2$. PCR products were separated by agarose gel electrophoresis (4% ultra high resolution agarose) and visualised by ethidiumbromide staining. The results from the PCR experiments are summarised in

Table 1. which also shows the sequence of PCR primers. primer annealing temperature (Tm). PCR product length and chromosome location of SSLP (with respect to the RI map of May 1997, Weeds World 4i).

F. Production of hybrid plants

5   *C24* plants heterozygous for *chlorina3A:Ac/GUS* are crossed as male to emasculated wildtype *Ler* to produce *Ler/C24(chlorina3A, GUS)* hybrid seeds.

Due to the heterozygosity of the *C24* parent, only 50 % of hybrid plants have inherited the *chlorina3A:Ac/GUS* locus. The remaining 50% of hybrid plants are wildtype with respect to *chlorina3A:Ac/GUS*. Since the mutant locus is linked to a kanamycin 10 resistance gene (contained on the same T-DNA as *GUS*) mutant plants can be pre-selected by germinating hybrid seeds on germination medium containing 50 mg/L kanamycin.

Ler plants homozygous for the five chromosome markers are male sterile *(ms1-1)* and are crossed without emasculation to wildtype *C24* to produce *Ler(an-1, py-1, gl1-1, cer2-1, ms1-1)/C24* hybrid seeds.

15   Other *Ler* plants. which are male fertile. are crossed after emasculation of the female parent to produce *Ler/C24* hybrid seeds.

G. Introduction of *MSH3* and *MSH6/3* antisense genes into Arabidopsis and analysis of meiotic homeologous recombination

(i) Transformation of hybrid plants and analysis of homeologous meiotic recombination

20   The plant transformation vectors comprising the antisense genes described in (B) and (C) above are introduced into *Agrobacterium tumefaciens* strain *AGL1* (Lazo et al.. 1991, Bio/Technology 9, 963-967) by electroporation. Recombinant *Agrobacterium* clones are selected on LB medium containing 50 mg/L rifampicin and 100 mg/L carbenicillin. Selected clones are used to infect roots of *Arabidopsis* hybrid plants (described in (F) 25 above) using the root transformation protocol of Valvekens et al. (1988, PNAS 85. 5536-5540) except that shoot and root inducing media contain hygromycin (10 mg/L) instead of kanamycin.

Plants regenerated from roots of hybrid plants are genetic clones of root donating plants and therefore are again genetic hybrids of two *Arabidopsis* subspecies described in 30 (F). However, in contrast to the root donating plants. the regenerated hybrid plants also contain the introduced transgene and the co-introduced hygromycin resistance gene with the latter allowing these plants to grow on hygromycin containing culture medium.

Hygromycin resistant plants are then allowed to enter the reproductive phase and to produce gametes by meiotic divisions of microspore and megaspore mothercells. At 35 meiosis, the *DMC1* promoter is activated and can direct the expression of antisense genes described in (B) and (C) above. leading to decreased *MSH6* and/or *MSH3* gene

expression. This in turn depletes the gamete mothercells of MSH6 and/or MSH3 protein, thus causing alteration of MMR during meiotic divisions and increasing the recombination frequency between homeologous chromosomes.

Transgenic plants are then allowed to self-fertilise and to produce seeds. These seeds (F2 seeds with respect to hybrid production), and the plants derived therefrom, carry the homeologous recombination events which can be identified by using the visual and molecular recombination markers described in (E) above.

In case of homeologous recombination between chromosomes of *Ler* and *C24(chlorina3A:Ac, GUS)*, the analysis concentrates on chromosome 2 by selecting plants showing the visual phenotypic marker *chlorina*. This marker thus serves as a reference point as it indicates that respective chromosomes 2 originate from *C24*. Other markers, such as *GUS* or molecular markers, on chromosome 2 may then be used to identify chromosomal regions which are derived from the *Ler* chromosome as a result of homeologous recombination. F2 plants of control transformants not expressing the antisense gene(s) can be analysed in parallel and the results can be used for comparison to homeologous recombination results obtained in antisense plants.

(ii) Transformation of C24 wildtype, hybrid plant production and analysis of homeologous meiotic recombination

Introduction of MMR altering genes into wildtype *C24* is done using the root transformation protocol as described in G(i) for transformation of hybrid plants. Transformed plants are selected by resistance to either 10 mg/L hygromycin (in case of transformation with T-DNA's derived from pNOS-Hyg-SCV) or to 50 mg/L kanamycin (in case of transformation with T-DNA's derived from pBIN19).

Transgenic plants are then allowed to self-fertilise and to produce seeds (T1 seeds). Segregation of the antibiotic resistance gene in the T1 population then indicates the number of transgene loci. Lines with a single transgene locus (indicated by a 3:1 ratio of resistant:sensitive plants) are selected and are bred to homozygosity. This is done by collecting selfed seeds (T2) from T1 plants and subsequent testing of at least four independent T2 seed populations for segregation of the antibiotic resistance gene. T2 populations which do not segregate the antibiotic resistance gene are assumed to be homozygous for both the resistance gene and the linked MMR altering gene.

C24 plants homozygous for the MMR altering gene are then crossed to *Ler* lines homozygous for recessive visual markers (see E(ii)) to produce *C24/Ler* hybrid plants as described in (F). These F1 hybrids are then allowed to enter the reproductive phase and to produce gametes by meiotic division of microspore and megaspore mothercells. At meiosis, the *DMC* 1 promoter is activated and can direct the expression of antisense or sense genes described in (B), (C) and (D) above, leading to decreased *MSH6* and/or *MSH3* gene expression. This in turn depletes the gamete mothercells of *MSH6* and/or *MSH3*

protein, thus causing alteration of MMR during meiotic divisions and increasing the recombination frequency between the homeologous chromosomes of *C24* and *Ler*. Recombination events are then scored in the F2 generation.

For recombination analysis, the hybrid plants are allowed to self-fertilise and to
5 produce F2 seeds. F2 plants are then preselected for a first visual marker. Since this marker is recessive, its visual presence indicates homozygosity for *Ler* DNA at the relevant locus. Those F2 plants which show this first visual marker are then analysed for the presence or absence of a second visual marker which in the *Ler* parent is closely linked to the first marker. Absence of the second visual marker indicates recombination between
10 the relevant *C24* and *Ler* chromosomes between the first and second marker. The frequency of recombination in transgenic hybrids is compared to the recombination frequency in control hybrids not expressing the MMR altering gene.

Examples of recombination analysis are the following.

The *Ler* line NW22(*dis1. ga4 . th1*) is used for crosses to transformed *C24*.

15 F2 plants are preselected first for thiamine requirement (*th1*) and then are further analysed for re-appearance of wildtype height (loss of *ga4*) and/or re-appearance of normal trichomes (loss of *dis1*) as a result of recombination.

The *Ler* line NW10(*tz-201, cer3*) is used for crosses to transformed *C24*.

F2 plants are then preselected first for thiazole requirement (*tz*) and then are further
20 analysed for re-appearance of normal. i.e. non-shiny stems (loss of *cer3*) as a result of recombination.

The *Ler* line NW134 (*ttg , ga3*) is used for crosses to transformed *C24*. F2 plants are first preselected for dwarfish appearance (*ga3*) and are then analysed for re-appearance of trichomes (loss of *ttg*) as a result of recombination.

25 *Ler* plants homozygous for *abi3*-1 and *gl1*-1 are used for crosses to transformed *C24*. F2 plants are first preselected for their ability to germinate in the presence of abscisic acid and are then analysed for re-appearance of trichomes on the leaves (loss of *gl1*-1) as a result of recombination.

In the case of homeologous recombination between transformed *C24* and the *Ler* line
30 NW1 (*an-1, py-1, gl1*-1, *cer2*-1, *ms1*-1), recombination analysis is similar the one described above, except that the second marker is not a visual marker but has to be a molecular marker. This is because the *Ler* parent carries only one visual marker per chromosome.

TABLE 1: SSLP Markers in *Arabidopsis thaliana* Subspecies

Chromosome	RI Map Position	PCR Primer Pair	Primer Sequence	Tm [°C]	length/COL	length/LER	length/C24
I	2.3	ATEAT1 F	GCCACTGCCTGAATGATATG	57.8	172	162	162
		ATEAT1 R	CGAACAGCCAACATTAATTCCC	58.2			
I	9.3	NGA63 F	AACCAAGGCACAGAAGCG	57.3	111	89	120
		NGA63 R	ACCCAAGTGATCGCCACC	59.6			
I	40.1	NGA248 F	TACCGAACCAAAACACAAAGG	56.1	143	129	no amplific.
		NGA248 R	TCTGTATCTCGGTGAATTCTCC	58.2			
I	81.2	NGA128 F	GGTCTGTTGATGTCGTAAGTCG	60.1	180	190	no amplific.
		NGA128 R	ATCTTGAAACCTTTAGGGAGGG	58.2			
I	81.2	NGA280 F	CTGATCTCACGGACAATAGTGC	60.1	105	85	85
		NGA280 R	GGCTCCATAAAAAGTGCACC	57.8			
I	111.4	NGA111 F	CTCCAGTTGGAAGCTAAAGGG	60	128	162	170
		NGA111 R	TGTTTTTAGGACAAATGGCG	70			
II	ca. 7.5	NGA168 F	CCTTCACATCCAAAACCCAC	57.8	213	217	208
		NGA168 R	GCACATACCCACCAACCAGAA	57.8			

II	ca. 48	NGA1126L	CGCTACGCTTTCGGTAAAG	57.8	191	199	196
		NGA1126R	GCACAGTCCAAGTCACAACC	59.9			
II	62.2	NGA361L	AAAGAGATGAGAATTTGGAC	51.7	114	120	114
		NGA361R	ACATATCAATATATTAAAGTAGC	49.5			
II	73	NGA168 F	TCGTCTACTGCACTGCCG	59.6	151	135	135
		NGA168 R	GAGGACATGTATAGGAGCCTCG	61.9			
II	ca. 77	AthBIO2 L	TGACCTCCTCTTCCATGGAG	59.9	141	209	139
		AthBIO2 R	TTAACAGAAACCCAAAGCTTTC	54.5			
II	ca. 83	AthUBIQUE L	AGGCAAATGTCCATTTCATTG	54.1	146	148	148
		AthUBIQUE R	ACGACATGGCAGATTTCTCC	57.8			
III	3.4	NGA172 F	AGCTGCTTCCTTATAGCGTCC	60	162	136	140
		NGA172 R	CATCCGAATGCCATTGTTC	55.4			
III	12.8	NGA126 F	GAAAAAACGCTACTTTCGTGG	56.1	119	147	no amplific.
		NGA126 R	CAAGAGCAATATCAAGAGCAGC	58.2			
III	17.5	NGA162 F	CATGCAATTTGCATCTGAGG	55.8	107	89	no amplific.
		NGA162 R	CTCTGTCACTCTTTCCTCTGG	60.1			

				Tm			
III	81.8	NGA6 F	TGGATTTCTTCCTCTCTTCAC	56.1	143	123	143
		NGA6 R	ATGGAGAAGCTTACACTGATC	56.1			
IV	19.8	NGA12 F	AATGTTGTCCTCCCCTCCTC	59.9	247	234	220
		NGA12 R	TGATGCTCTCTGAAACAAGAGC	58.2			
IV	24.1	NGA8 F	GAGGGCAAATCTTTATTTCGG	56.1	154	198	190
		NGA8 R	TGGCTTTCGTTTATAAACATCC	54.5			
IV	102	NGA1107 L	GCGAAAAACAAAAAAATCCA	50.2	150	140	140
		NGA1107 R	CGACGAATCGACAGAATTAGG	58			
V	11.8	NGA225 F	GAAATCCAAATCCCAGAGAGG	58	119	189	119
		NGA225 R	TCTCCCCACTAGTTTTGTGTCC	60.1			
V	16.7	NGA249 F	TACCGTCAATTTCATGCC	55.4	125	115	115
		NGA249 R	GGATCCCTAACTGTAAAATCCC	58.2			
V	19.9	CA72 F	AATCCCAGTAACCAAACACACA	56.3	124	110	110
		CA72 R	CCCAGTCTAACCACGACCAC	61.9			
V	20	NGA151 F	GTTTTGGGAAGTTTTGCTGG	55.8	150	120	130
		NGA151 R	CAGTCTAAAAGCGGAGAGTATGATG	58.6			

>							
>	24	NGA106 F	GTTATGGAGTTTCTAGGGCACG	60.1	157	123	130
		NGA106 R	TGCCCCATTTTGTTCTTCTC	55.8			
>	37.8	NGA139 F	AGAGCTACCAGATCCGATGG	59.9	174	132	132
		NGA139 R	GGTTTCGTTTCACTATCCAGG	55.8			
>	50	NGA76 F	GGAGAAAATGTCACTCTCCACC	60.1	231	>250	300
		NGA76 R	AGGCATGGGAGACATTTACG	57.8			
>	61.1	ATHSO191 L	CTCCACCAATCATGCAAATG	55.8	148	156	146
		ATHSO191 R	TGATGTTGATGGAGATGGTCA	53.7			
>	81.7	NGA129 F	TCAGGAGGAACTAAAGTGAGGG	60.1	177	179	172
		NGA129 R	CACACTGAAGAGATGGTCTTGAGG	60.1			

CLAIMS

1.   An isolated and purified DNA molecule comprising a polynucleotide sequence encoding a polypeptide functionally involved in the DNA mismatch repair system of a plant.

5   2.   A DNA molecule according to claim 1 wherein said polypeptide is homologous to a mismatch repair polypeptide of a yeast or of a human.

3.   A DNA molecule according to claim 1 wherein said polypeptide is homologous to AtMSH3 (SEQ ID NO: 19) or to AtMSH6 (SEQ ID NO: 31).

4.   An isolated and purified polypeptide functionally involved in the DNA
10 mismatch repair system of a plant.

5.   A polypeptide according to claim 4 which is homologous to a mismatch repair polypeptide of a yeast or of a human.

6.   An isolated and purified polypeptide selected from the group consisting of a polypeptide encoded by the gene *AtMSH3* (SEQ ID NO: 18), a polypeptide encoded by the
15 gene *AtMSH6* (SEQ ID NO:30), polypeptides homologous to a polypeptide encoded by the gene *AtMSH3* (SEQ ID NO: 18) and polypeptides homologous to a polypeptide encoded by the gene *AtMSH6* (SEQ ID NO:30).

7.   An isolated and purified DNA molecule comprising a polynucleotide sequence selected from the group consisting of (i) a sequence encoding a polynucleotide which is
20 capable of interfering with the expression of a plant polynucleotide sequence encoding a polypeptide which is homologous to a mismatch repair polypeptide of a yeast or of a human and thereby disabling said plant polynucleotide sequence; and (ii) a sequence encoding a polypeptide capable of disrupting the DNA mismatch repair system of a plant.

8.   A DNA molecule according to claim 7 comprising a polynucleotide sequence
25 encoding a polynucleotide capable of interfering with the expression of a plant polynucleotide sequence encoding a polypeptide which is homologous to a mismatch repair polypeptide of a yeast or of a human and thereby disabling said plant polynucleotide sequence.

9.   A DNA molecule according to claim 8 wherein said polynucleotide is capable
30 of interfering with the expression of a plant polynucleotide sequence is a sense polynucleotide, an antisense polynucleotide or a ribozyme.

10.   A DNA molecule according to claim 7 comprising a polynucleotide sequence encoding a polypeptide capable of disrupting the DNA mismatch repair system of a plant.

11. A DNA molecule according to claim 10 wherein said polypeptide is homologous to AtMSH3 (SEQ ID NO: 19) or to AtMSH6 (SEQ ID NO: 31).

12. A DNA molecule according to claim 10 further comprising a regulation element capable of causing overexpression of said polypeptide in a cell of said plant.

13. A chimeric gene comprising:

a DNA sequence selected from the group consisting of (i) a sequence encoding a polynucleotide capable of interfering with the expression of a plant polynucleotide sequence encoding a polypeptide which is homologous to a mismatch repair polypeptide of a yeast or of a human and thereby disabling said plant polynucleotide sequence, and (ii) a sequence encoding a polypeptide capable of disrupting the DNA mismatch repair system of a plant; and

at least one regulation element capable of functioning in a plant cell.

14. A chimeric gene according to claim 13 wherein said regulation element is selected from constitutive, inducible, tissue type specific and cell type specific promoters.

15. A chimeric gene according to claim 13 comprising a DNA sequence encoding a polypeptide capable of disrupting the DNA mismatch repair system of a plant, wherein said regulation element is capable of causing overexpression of said polypeptide in a cell of said plant.

16. A chimeric gene according to claim 13 wherein said regulation element is selected from the group consisting of 35S, NOS, PR1a, AoPR1 and DMC1.

17. A plasmid or vector comprising a chimeric gene according to any one of claims 13-16.

18. A plant cell stably transformed, transfected or electroporated with a plasmid or vector according to claim 17.

19. A plant comprising a cell according to claim 18.

20. A plant according to claim 19 selected from plants of the families *Brassicaceae, Poaceae, Solanaceae, Asteraceae, Malvaceae, Fabaceae, Linaceae, Canabinaceae, Dauaceae* and *Cucurbitaceae*.

21. A process for at least partially inactivating a DNA mismatch repair system of a plant cell, comprising transforming or transfecting said plant cell with a DNA molecule according to any one of claims 1-3 or 7-12 and causing said DNA sequence to express said polynucleotide or said polypeptide.

22. A process for at least partially inactivating a DNA mismatch repair system of a plant cell, comprising transforming or transfecting said plant cell with a chimeric gene

according to any one of claims 13-16 and causing said DNA sequence to express said polynucleotide or said polypeptide.

23. A process for at least partially inactivating a DNA mismatch repair system of a plant cell, comprising transforming or transfecting said plant cell with a plasmid or vector according to claim 17 and causing said DNA sequence to express said polynucleotide or said polypeptide.

24. A process for increasing genetic variation in a plant comprising obtaining a hybrid plant from a first plant and a second plant, or cells thereof, said first and second plants being genetically different; altering the mismatch repair system in said hybrid plant; permitting said hybrid plant to self-fertilise and produce offspring plants; and screening said offspring plants for plants in which homeologous recombination has occurred.

25. A process according to claim 24 wherein a first gene is incapacitated in said first plant, a second gene is incapacitated in said second plant, and said first and second genes are incapacitated in said hybrid plant thereby altering the mismatch repair system of said hybrid plant.

25. A process according to claim 25 wherein said incapacitation of the mismatch repair system of said hybrid plant is reversible.

26. A process according to claim 24 wherein a new genetic linkage of a desired characteristic trait or of a gene which contributes to a desired characteristic trait is observable in at least one of said offspring plants.

27. A process for obtaining a plant having a desired characteristic, comprising altering the mismatch repair system in a plant, cell or plurality of cells of a plant which does not have said desired characteristic, permitting mutations to persist in said cells to produce mutated plant cells, deriving plants from said mutated plant cells, and screening said plants for a plant having said desired characteristic.

28. A process according to claim 27 wherein said step of altering the mismatch repair system comprises introducing into said hybrid plant, plant, cell or cells a chimeric gene according to claim 13 and permitting the chimeric gene to express a polynucleotide which is capable of interfering with the expression of a plant polynucleotide sequence in a mismatch repair gene of the hybrid plant, plant, cell or cells, or a polypeptide capable of disrupting the DNA mismatch repair system of the hybrid plant, cell or cells.

29. A process according to claim 28 comprising inactivating an MSH3 gene and/or an MSH6 gene of said plant.

30. A process according to claim 28 comprising inactivating an MSH3 gene and an MSH6 gene of said plant.

31. A process according to claim 27 comprising at least partially inactivating the mismatch repair system of said plant in a predetermined cell type or in a predetermined tissue of said plant.

32. A process according to claim 31 further comprising restoring mismatch repair in said cell type or said tissue.

33. An oligonucleotide capable of hybridising at 45°C under standard PCR conditions to a DNA molecule according to claim 1 with the proviso that said oligonucleotide is other than SEQ ID NO:1 or SEQ ID NO:2.

34. An oligonucleotide capable of hybridising at 45°C under standard PCR conditions to the DNA of SEQ ID NO: 18 with the proviso that said oligonucleotide is other than SEQ ID NO:1 or SEQ ID NO:2.

35. An oligonucleotide capable of hybridising at 45°C under standard PCR conditions to the DNA of SEQ ID NO:30 with the proviso that said oligonucleotide is other than SEQ ID NO:1 or SEQ ID NO:2.

Figure 1
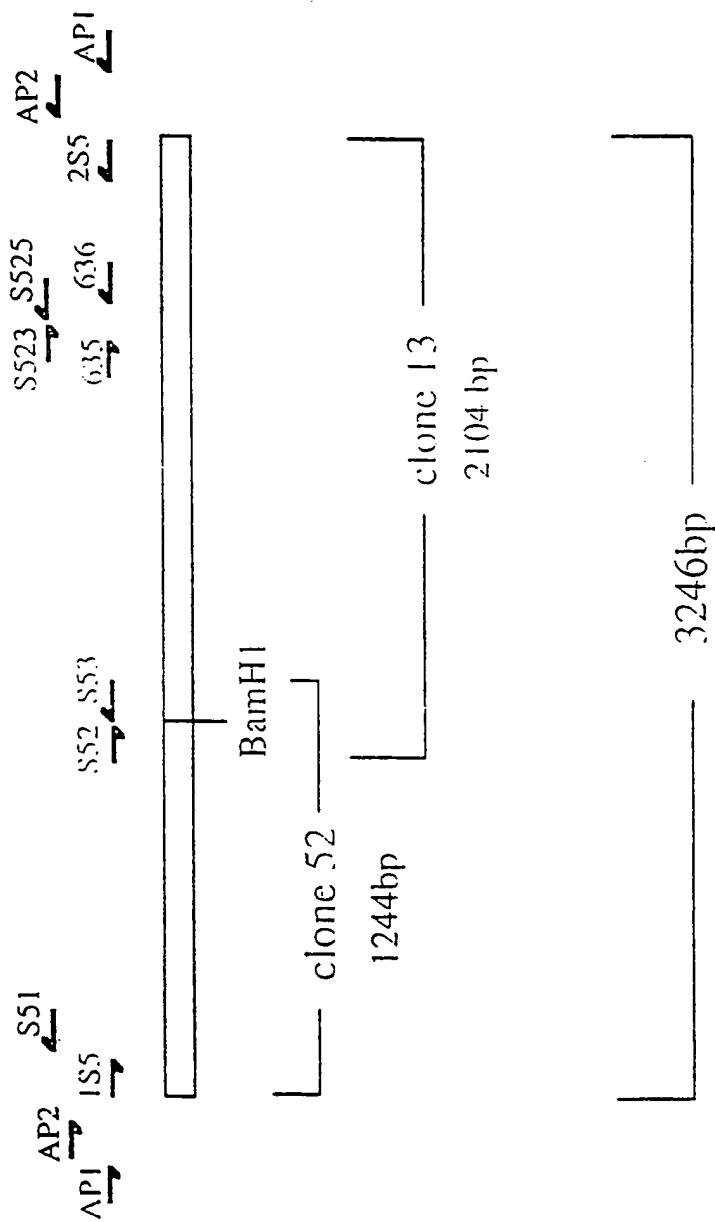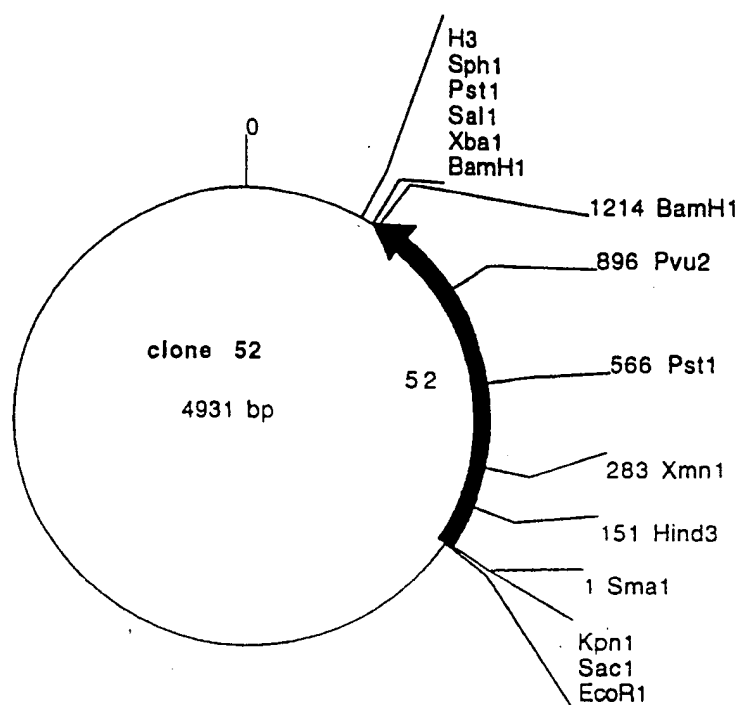
Figure 2

Comments/References: 52= 3' side of S5 (AtMSH3) 1244bp in pUC18/Sma1
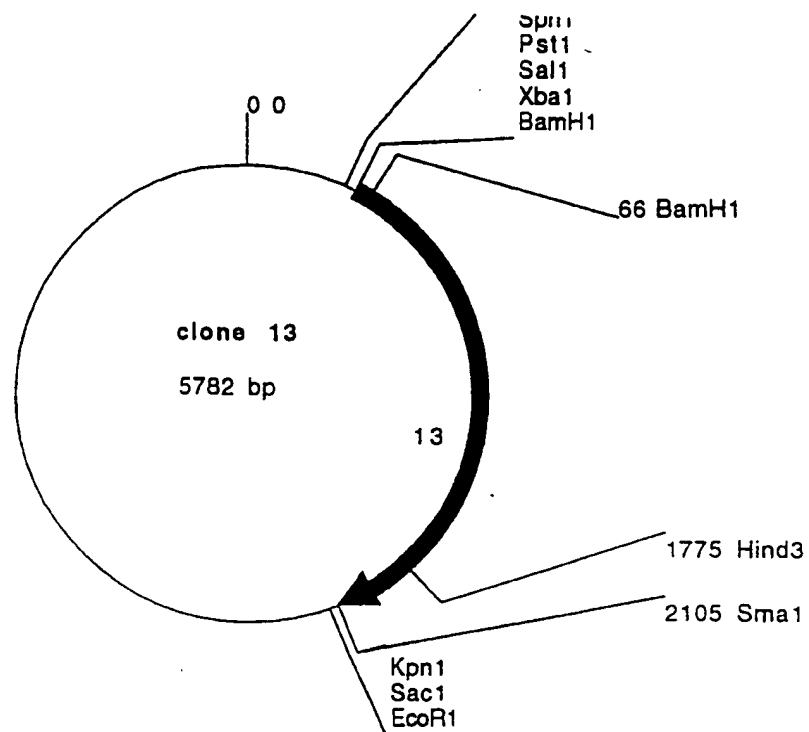
Figure 3

Comments/References: 13 = 3' side of S5 (AtMSH3) 2104bp in pUC18/Sma1

```
  1    CCTAAGAAAGCGCGCGAAAATTGGCAACCCAAGTTCGCCATAGCCACGACCACGACCTTCCATTCTCTTAAACGGAGGA      80

 81    GATTACGAATAAAGCAATT ATG GGC AAG CAG CAG ACG ATT TCT TTC TTC GCT CCC                    144
  1                        M   G   K   Q   Q   T   I   S   R   F   F   A   P                  15

145    AAA CCC AAA TCC CCG ACT CAC GAA AAT CCG GTA GCC GAA TCA ACA CCG CCG                    204
 16    K   P   K   S   P   T   H   E   N   P   V   A   E   S   T   P   P                      35

205    AAG ATA TCC GCC ACT GTA TCC TTC TCT CCT TCC AAG CTT CTC TCC GAC CAC CTC                264
 36    K   I   S   A   T   V   S   F   S   P   S   K   L   L   S   D   H   L                  55

265    GCC GCC GCG TCA CCC AAA AAG CCT AAA CTT TCT CCT CAC CAA AAC CCA GTA CCC GAT            324
 56    A   A   A   S   P   K   K   P   K   L   S   P   H   Q   N   P   V   P   D              75

325    CCC AAT TTA CAC CAA AGA TTT CTC CAG AGA TTT CTG GAA CCC TCG CCG GAG GAA TAT GTT        384
 76    P   N   L   H   Q   R   F   L   Q   R   F   L   E   P   S   P   E   E   Y   V          95

385    CCC GAA ACG TCA TCG AGG AAA TAC ACA CCA GAA CAG CAA CAG GAG CTA AAG                    444
 96    P   E   T   S   S   R   K   Y   T   P   E   Q   Q   Q   E   L   K                      115

445    AGC AAG TAC TCA GAT GTG TAC GGT GAA GTT TAC AGG TTC TTC GGA                            504
116    S   K   Y   S   D   V   Y   G   E   V   Y   R   F   F   G                              135

505    GAA GAC GCG ATC GCA GCA CGC CGC GTT TTG GGT ATT TAC GCT CAT ATG GAT CAC AAT TTC        564
136    E   D   A   I   A   A   R   R   V   L   G   I   Y   A   H   M   D   H   N   F          155

565    ATG ACG GCG AGT GTG CCA ACA TTT CGA AAT TTC CAT GTG AGA AGA CTG GTG AAT GCA            624
156    M   T   A   S   V   P   T   F   R   N   F   H   V   R   R   L   V   N   A              175

625    GGA TAC AAG ATT GGT GTA GTG AAG CAG ACT GAA GCC ATT AAG TCC CAT GGT GCA                684
176    G   Y   K   I   G   V   V   K   Q   T   E   A   I   K   S   H   G   A                  195

665    AAC CGG ACC GGC CCT TTT TTC CGG CTG GCG TCG GCG TTG TAT ACC AAA GCC ACG CTT GAA        744
196    N   R   T   G   P   F   F   R   L   A   S   A   L   Y   T   K   A   T   L   E          215

745    GCG GCT GAG GAT ATA AGT GTG GAT GAG GGA TTT GGT GGT TCA CAG AGT AAT                    804
216    A   A   E   D   I   S   V   D   E   G   F   G   G   S   Q   S   N                      235

805    TTC TTG GTT TGT GTT GTG GAT GAG AGA GTT AAG TCG GAG ACA TTA GGC ATT GAA                864
236    F   L   V   C   V   V   D   E   R   V   K   S   E   T   L   G   I   E                  255

865    ATG AGT TTT GAT GTT AGA GTC GGT GTT GTT GGC GAA ATT TCG ACA GGT GAA GTT                924
256    M   S   F   D   V   R   V   G   V   V   G   E   I   S   T   G   E   V                  275
```

Figure 4

```
 925  TAT GAA GAG TTC AAT GAT AAT ATG AGA AGT GGA TTA GAG GCT GTG ATT TTG AGC TTG   984
 276   Y   E   E   F   N   D   N   M   R   S   G   L   E   A   V   I   L   S   L    295

 985  TCA CCA GCT GAG CTG TTG GGC CAG CCT CTT TCA GAG CAA CAA ACT GAG AAG TTT GTG  1044
 296   S   P   A   E   L   L   G   Q   P   L   S   E   Q   Q   T   E   K   F   V    315

1045  GCA CAT GCT GGA CCT ACC TCA AAC GTT CGA GTG GAA CGT GAT TGT TTC AGC          1104
 316   A   M   A   G   P   T   S   N   V   R   V   E   R   D   C   F   S            335

1105  AAT GGT GCA GTA GAT GAG GTT ATT TCA GAT GCA AAA ATC AGC GGT AAC TTA          1164
 336   N   G   A   V   D   E   V   I   S   D   A   K   I   S   G   N   L            355

1165  GAA GAT GAT AAA GAA ATG AAG CTG GAG GCT GCA GAA AAA GGA TGC TTG ACA GTT      1224
 356   E   D   D   K   E   M   K   L   E   A   A   E   K   G   C   L   T   V        375

1225  CAT ACA ATT ATG AAC ATG CCA CAT CTG ACT GTT CAA ACG TTT CAT                  1284
 376   H   T   I   M   N   M   P   H   L   T   V   Q   T   F   H                    395

1285  CTC AAA CAG TTT GGA TTT GAA GAG AGG ATC CTC TAC CGC TCT TTG TCA              1344
 396   L   K   Q   F   G   F   E   E   R   I   L   Y   R   S   L   S                415

1345  AGT AAC ACA GAG ATG AAC ACT CTC TCA GCC AAT GTT GTG AAA AAT                  1404
 416   S   N   T   E   M   N   T   L   S   A   N   V   V   K   N                    435

1405  AAT TCA GAT GGA TCG GAA ATG AAT CAC AAT CAC ACA CTT GTA                      1464
 436   N   S   D   G   S   E   M   N   H   N   H   T   L   V                        455

1465  TAT GCT TCC AGG CTT CTT AGA CAC TGG CTA TGC GAT AGA AAT TTG ATA              1524
 456   Y   A   S   R   L   L   R   H   W   L   C   D   R   N   L   I                475

1525  TCT GCT CGG CTT GAT GCT GTT GAG ATT TCT GGA TCT CAT AGT TCC                  1584
 476   S   A   R   L   D   A   V   E   I   S   G   S   H   S   S                    495

1585  CAG CTC AGC AGT GAG TTG GTT GAA GGT GAA GCA GTA TCA CCT GAG TTT              1644
 496   Q   L   S   S   E   L   V   E   G   E   A   V   S   P   E   F                515

1645  TAT CTC GTG CTC TCA GTC TTG ACA GCT AAA GAG TTC ATT CAA CGT GGA              1704
 516   Y   L   V   L   S   V   L   T   A   K   E   F   I   Q   R   G                535

1705  ATA ACA AGA ATC TTT CAT CGG ACT GCT ATG TCT GAT ATT GCA GTT ATG GAA GCT      1764
 536   I   T   R   I   F   H   R   T   A   M   S   D   I   A   V   M   E   A        555

1765  ATT TTA CTT GCG GGG AAG CAA ATT CAG CGG CTT GGC ATA AAG CAA GAC TCT GAA ATG AGG  1824
 556   I   L   L   A   G   K   Q   I   Q   R   L   G   I   K   Q   D   S   E   M   R    575
```

Figure 4 (Continued)

```
1825                                                                                      1884
 576  AGT ATG CAA TCT GCA ACT GTG CGA TCT ACT CTT TTG AGA AAA TTG ATT TCT GTT ATT TCA      595
       S   M   Q   S   A   T   V   R   S   T   L   L   R   K   L   I   S   V   I   S

1885                                                                                      1944
 596  TCC GTT GTG GAC AAT GCC GGA AAA CTT CTC TCT GCC CTA AAT AAG GAA GCG GCT              615
       S   V   V   D   N   A   G   K   L   L   S   A   L   N   K   E   A   A

1945                                                                                      2004
 616  GTT CGA GGT GAC TTG CTC GAC ATA CTA ATC ACT TCC AGC GAC CAA TTT CCT GAG CTT GCT      635
       V   R   G   D   L   L   D   I   L   I   T   S   S   D   Q   F   P   E   L   A

2005                                                                                      2064
 636  GAA GCT CGC CAA GCA GTT TTA GTC ATC AGG GAA AAG CTG GAT TCC ATC AGG TCA TTT          655
       E   A   R   Q   A   V   L   V   I   R   E   K   L   D   S   I   R   S   F

2065                                                                                      2124
 656  CGC AAG AAG CTC ATT GCT CGA TTT CTT CAA GTG TCG GGG ATC ACA CAT TTG                  675
       R   K   K   L   I   A   R   F   L   Q   V   S   G   I   T   H   L

2125                                                                                      2184
 676  ATA GAG CTG CCC GTT GAT TCC AAG GTC CCT ATG AAT TGG GTG AAA GTA AAT AGC ACC AAG      695
       I   E   L   P   V   D   S   K   V   P   M   N   W   V   K   V   N   S   T   K

2185                                                                                      2244
 696  AAG ACT ATT CGA TAT CAT CCC CCA GAA ATA GTA GCT GGC TTG GAT GAG CTA GCA              715
       K   T   I   R   Y   H   P   P   E   I   V   A   G   L   D   E   L   A

2245                                                                                      2304
 716  ACT GAA CAT CTT GCC ATT GTG AAC CGA GCT TCG GAT AGT TTC CTC AAG AGT                  735
       T   E   H   L   A   I   V   N   R   A   S   D   S   F   L   K   S

2305                                                                                      2364
 736  AGA TAC TAC ACA GAT TTT AAG GCT GCC GTT CAA GCA CTG GCA GAC TGT TTG CAC              755
       R   Y   Y   T   D   F   K   A   A   V   Q   A   L   A   D   C   L   H

2365                                                                                      2424
 756  TCC CTT TCA ACT CTA TCT AGA AAC TAT AAG AAC CGA GAG CCC CGT GTA CTG GTG GAT GAC      775
       S   L   S   T   L   S   R   N   Y   K   N   R   E   P   R   V   L   V   D   D

2425                                                                                      2484
 776  GAA CCA CCA GAG ATA CAG TCT ACA AAT GAC ACA GAA GGG GAA TAT GTC GAG ACT ATA TTA      795
       E   P   P   E   I   Q   S   T   N   D   T   E   G   E   Y   V   E   T   I   L

2485                                                                                      2544
 796  GAT AAC TTC GTC CCA ATT TTG CAT GCA GAA GGG GAA TAT TGC CAA GTT GAT GAC             815
       D   N   F   V   P   I   L   H   A   E       G   E   Y   C   Q   V   D   D

2545                                                                                      2604
 816  ACC GGA CCT AAC ATG GGA GGA AAG AGC TGC TAT ATC CGT CAA GTT GCC AAG CTG CAC          835
       T   G   P   N   M   G   G   K   S   C   Y   I   R   Q   V   A   K   L   H

2605                                                                                      2664
 836  ATG GCT CAG GTT CCA GTA CCA TCA GCG TTC GCC AAG CTG CAC GTG GAT GGT                  855
       M   A   Q   V   P   V   P   S   A   F   A   K   L   H   V   D   G

2665                                                                                      2724
 856  GTT TTC ACT CGG ATG GGT GCT TCA GAC AGT ATC CAG CAT GGC AGA AGT ACC TTT CTA GAA      875
       V   F   T   R   M   G   A   S   D   S   I   Q   H   G   R   S   T   F   L   E
```

Figure 4 (Continued)

```
2725                                                                              2784
876  GAA TTA AGT GAA GCG TCA CAC ATA ATC AGA ACC TGT TCT TCT TCG CTT GTT ATA TTA  895
      E   L   S   E   A   S   H   I   I   R   T   C   S   S   S   L   V   I   L

2785                                                                              2844
896  GAT GAG CTT GGA AGA GGC ACT AGC ACA CAC GAC GGT GTA GCC ATT GCC TAT GCA ACA TTA  915
      D   E   L   G   R   G   T   S   T   H   D   G   V   A   I   A   Y   A   T   L

2845                                                                              2904
916  CAG CAT CTC CTA GCA GAA AAG AGA AAC GGA TTC CCA CAT TAC CCT GAA ATA  935
      Q   H   L   L   A   E   K   R   N   G   F   P   H   Y   P   E   I

2905                                                                              2964
936  GCT GAG ATC AGT AAC GGA TTC CCA GGT TCT GTT GGG ACA TAC CAT GTC TCG TAT CTG ACA  955
      A   E   I   S   N   G   F   P   G   S   V   G   T   Y   H   V   S   Y   L   T

2965                                                                              3024
956  TTG CAG AAG GAT AAA GGC AGT TAT GAT CAT GAT GTG ACC TAC CTA TAT AAG CTT GTG  975
      L   Q   K   D   K   G   S   Y   D   H   D   V   T   Y   L   Y   K   L   V

3025                                                                              3084
976  CGT GGT CTT TGC AGC AGG AGC TTT GGT TTT AAG GTT GCT CTT CAG ATA CCT CCA  995
      R   G   L   C   S   R   S   F   G   F   K   V   A   L   Q   I   P   P

3085                                                                              3144
996  TCA TGT ATA CGT CGA GCC ATT TCA ATG GCT GAA AAA TTG GAA GTA CGT GCA AGA  1015
      S   C   I   R   R   A   I   S   M   A   E   K   L   E   V   R   A   R

3145                                                                              3204
1016 GAG AGA AAT ACA CGC ATG GGA GAA CCA GAA CAT GGC GCA GAA GGA GAA GAA  1035
      E   R   N   T   R   M   G   E   P   E   H   G   A   E   G   E   E

3205                                                                              3264
1036 TCT ATT TCG GCT CTA GGT GAC TTG TTT GCA GAC CTG AAA TTT GCT CTC TCT GAA GAG GAC  1055
      S   I   S   A   L   G   D   L   F   A   D   L   K   F   A   L   S   E   E   D

3265                                                                              3324
1056 CCT TGG AAA GCA TTC GAG TTT TTA AAG CAT GCT TGG AAG ATT GCT GGC AAA ATC AGA CTA  1075
      P   W   K   A   F   E   F   L   K   H   A   W   K   I   A   G   K   I   R   L

3325                                                                              3397
1076 AAA CCA ACT TGT TCA TTT TGA TTTAATCTTAACATTATAGCAACTGCAAGGTCTTGATCATCTGTTAGTTGCG  1082
      K   P   T   C   S   F   *

3398                                                                              3458
1    TACTAACTT ATG TGT ATT AGT ATA ACA AGA AAA GAG AAT TAG AGAG ATG GAT TCT AAT CCG  5
               M   C   I   S   I   T   R   K   E   N   *         M   D   S   N   P

3459                                                                              3522
6    GTG TTG CAG TAC ATC TTT TCT CCA CCC GCA TAA AAAAAAAAAAAAAAAAAAAAAAAA  16
      V   L   Q   Y   I   F   S   P   P   A   *
```

Figure 4 (Continued)

Figure 5
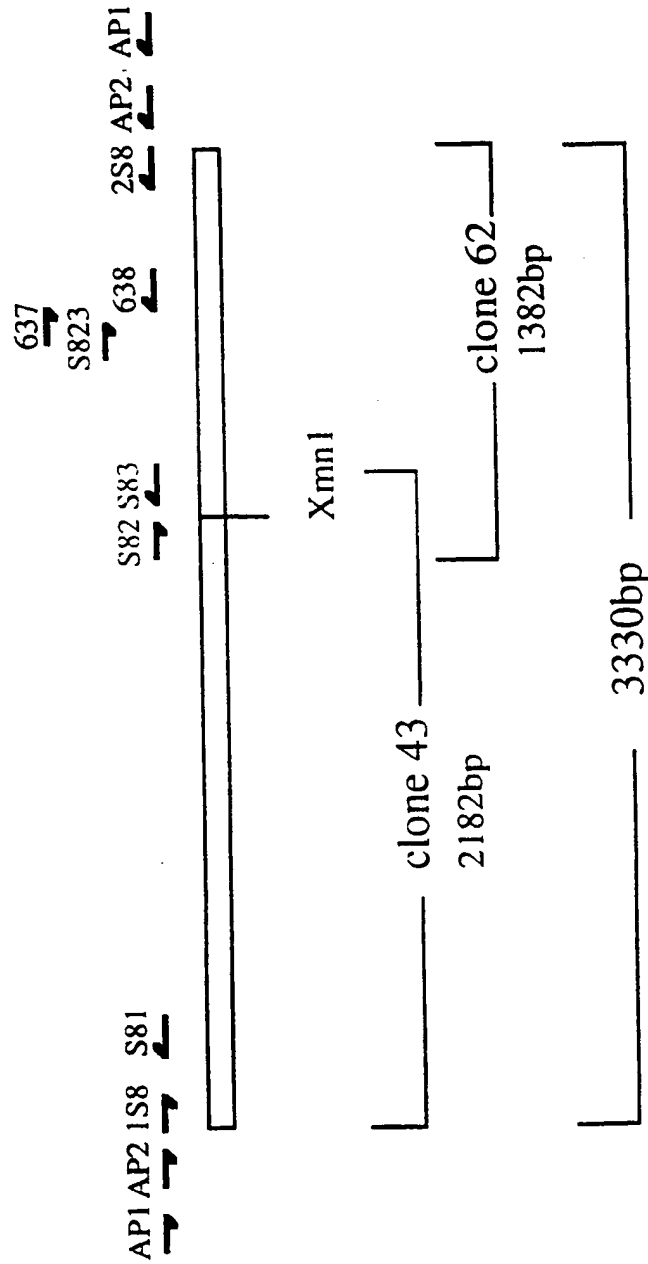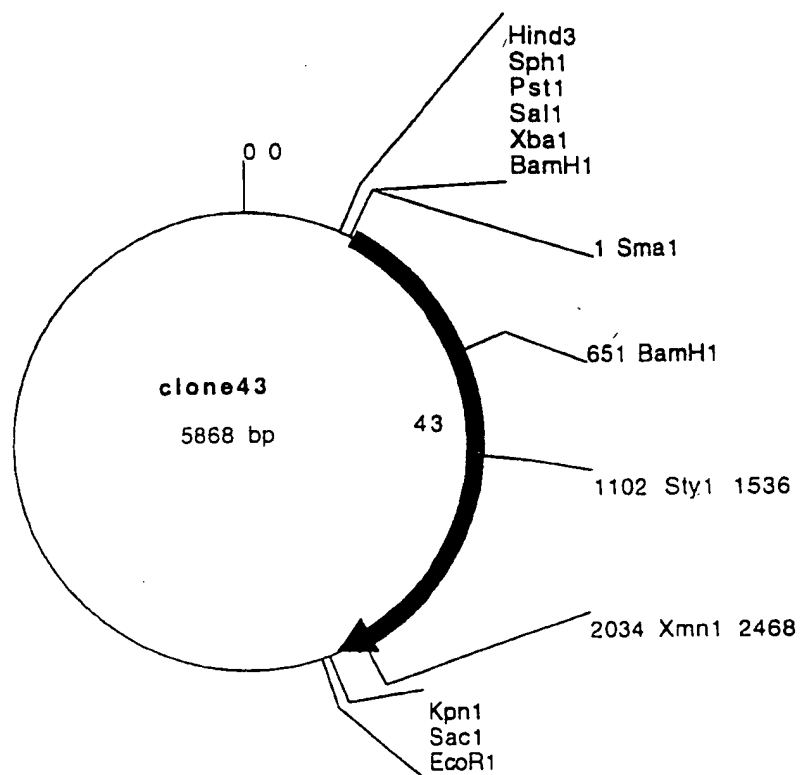
Figure 6

Figure 7

Comments/References: 43= 5' side of S8 (AtMSH6) 2182 bp in pUC18/Sma1
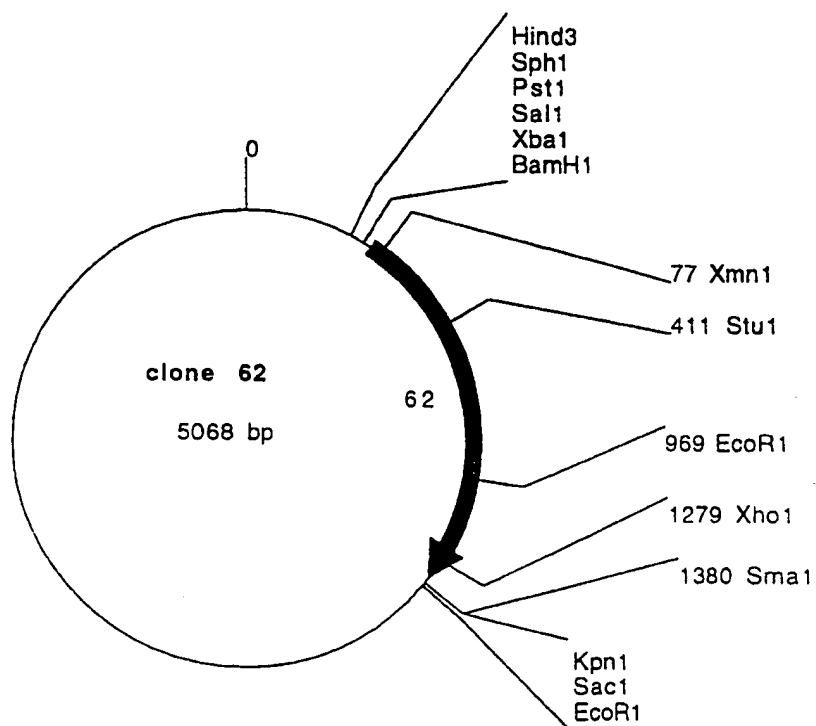
Figure 8

Comments/References: 62= 3' side of S8 (AtMSH6) 1379bp in pUC18/Sma1

```
1    AAAAGTTGAGCCCTGAGGAGTATCGTTTCCGCGCCATTCTACGACGCAAGGCGAAAATTTTTGGGCGCCAATCTTTCCCCCC    80

81   TTTCGAATTCTCAGCTCAAAACATCGTTTCTCTCTCTCACTCTCTCTCAAAA ATG CAG CGC CAG                 153
1                                                         M   Q   R   Q                   4

154  AGA TCG ATT TTG TCT TTC TTC CAA AAA CCC ACC GCG GCG ACT ACG AAG GGT TTG GTT TCC       213
5    R   S   I   L   S   F   F   Q   K   P   T   A   A   T   T   K   G   L   V   S          24

214  GGC GAT GCT GCT AGC GGG GGC GGA GGA CCA CGA TTT AAT GTG AAG GAA GGG                   273
25   G   D   A   A   S   G   G   G   G   P   R   F   N   V   K   E   G                      44

274  GAT GCT AAA GGC GAC GCT TCT GTA CGT TTT GCT GTT TCG AAA TCT GTC GAT GAG GTT AGA       333
45   D   A   K   G   D   A   S   V   R   F   A   V   S   K   S   V   D   E   V   R          64

334  GGA ACG GAT ACT CCA CCG GAG AAG GTT CCG CGT GTC CTG CCG TCT GGA TTT AAG CCG           393
65   G   T   D   T   P   P   E   K   V   P   R   V   L   P   S   G   F   K   P              84

394  GCT GAA TCC GCC GST GAT GCT TCG TCC AAT ATT ATG CAT AAG TTT GTA AAA                   453
85   A   E   S   A   G   D   A   S   S   N   I   M   H   K   F   V   K                      104

454  GTC GAT GAT CGA GAT TGT TCT GGA GAG AGC CGA GAA GAT GTT GTT CCG CTG AAT GAT           513
105  V   D   D   R   D   C   S   G   E   R   S   R   E   D   V   V   P   L   N   D          124

514  TCA TCT CTA TGT ATG AAG GCT AAT GAT GTT ATT CCT CAA TTT CGT TCC AAT AAT GGT AAA       573
125  S   S   L   C   M   K   A   N   D   V   I   P   Q   F   R   S   N   N   G   K          144

574  ACT CAA GAA AGA AAC CAT GCT TTT AGT GGG AGA GCT GAA CTT AGA TCA GTA GAA               633
145  T   Q   E   R   N   H   A   F   S   G   R   A   E   L   R   S   V   E                  164

634  GAT ATA GGA GTA GAT GGC GAT GTT CCT GGT CCA GAA ACA CCA ATG CGT CCA CGT GCT           693
165  D   I   G   V   D   G   D   V   P   G   P   E   T   P   M   R   P   R   A              184

694  TCT CGC TTG AAG CGA GTT CTG GAG GAT GAA ATG ACT TTT AAG GAG GAT AAG GTT CCT GTA       753
185  S   R   L   K   R   V   L   E   D   E   M   T   F   K   E   D   K   V   P   V          204

754  TTG GAC TCT AAC AGG CTG AAA ATG CTC CAG GAT CCG GTT TGT GGA GAG AAG AAA GAA           813
205  L   D   S   N   R   L   K   M   L   Q   D   P   V   C   G   E   K   K   E              224

814  GTA AAC GAA ACC AAA TTT GAA TGG CTT TCT CGA ATC AGG GAT GCC AAT AGA                   873
225  V   N   E   T   K   F   E   W   L   S   R   I   R   D   A   N   R                      244

874  AGA CGT CCT GAT GAT CCC CTT TAC GAT AGA AAG ACC TTA CAC ATA CCA CCT GAT GTT TTC       933
245  R   R   P   D   D   P   L   Y   D   R   K   T   L   H   I   P   P   D   V   F          264
```

Figure 9

```
934/265   AAG AAA ATG TCT GCA TCA CAA AAG CAA TAT TGG AGT GTT AAG AGT GAA TAT ATG GAC ATT   993/284
           K   K   M   S   A   S   Q   K   Q   Y   W   S   V   K   S   E   Y   M   D   I

996/285   GTG CTT TTC AAA GTG GGG AAA TTT TAT GAG CTG TAT GAG GCG GAA TTA GGT               1053/304
           V   L   F   K   V   G   K   F   Y   E   L   Y   E   A   E   L   G

1054/305  CAC AAG GAG CTT GAC TGG ATG AAG ACC ATG AGT ... CAG GTT GGT                        1113/324
           H   K   E   L   D   W   M   K   T   M   S       Q   V   G

1114/325  ATC TCT GAA AGT GGG ATA GAT GAG GCA GTG CAA AAG CTA TTA GCT CGT GGA TAT AAA GTT   1173/344
           I   S   E   S   G   I   D   E   A   V   Q   K   L   L   A   R   G   Y   K   V

1174/345  GGA CGA ATC GAG CAG CTA GAA ACA TCT GAC CAA GCA AAA GCC AGA GGT GCT AAT ACT ATA   1233/364
           G   R   I   E   Q   L   E   T   S   D   Q   A   K   A   R   G   A   N   T   I

1234/365  ATT CCA AGG AAG CTA GTT CAG GTA CTA ACT CCA TCA ACA GCA TCA GAG GGA AAC ATC GGG   1293/384
           I   P   R   K   L   V   Q   V   L   T   P   S   T   A   S   E   G   N   I   G

1294/385  CCT GAT GCC GTC CAT CTT CTT GCT ATA AAA GAG ATC AAA GAG CTA CAA AAG TGT TCA       1353/404
           P   D   A   V   H   L   L   A   I   K   E   I   K   E   L   Q   K   C   S

1354/405  ACT GTG TAT GGA TTT GCT TTT GTT GAC TGT GCC TTG AGG TTT TGG GTT GGG TCC ATC       1413/424
           T   V   Y   G   F   A   F   V   D   C   A   L   R   F   W   V   G   S   I

1414/425  AGC GAT GAT GCA TCA TGT GCT CTT GGA GCG TTA ATG CAG GTT TCT CCA AAG GAA           1473/444
           S   D   D   A   S   C   A   L   G   A   L   M   Q   V   S   P   K   E

1474/445  GTG TTA TAT TAT GAC AGT AAA GGG CTA TCA TCA AGA GAA GCA CAA AAG GCT CTA AAA TAT ACG   1533/464
           V   L   Y   Y   D   S   K   G   L   S   S   R   E   A   Q   K   A   L   K   Y   T

1534/465  TTG ACA GGG TCT ACG GCG GTA CAG CCA GTA TTG GCT CCA GTT GGG GAT ATG GGG GAT ACA GAT   1593/484
           L   T   G   S   T   A   V   Q   P   V   L   A   P   V   G   D   M   G   D   T   D

1594/485  GCT GCT GGA GTT AGA AAT ATA GAA TCT AAC GGA TAC TTT AAA GGT TCT TCT GAA TCA       1653/504
           A   A   G   V   R   N   I   E   S   N   G   Y   F   K   G   S   S   E   S

1654/505  TGG AAC TGT GAT GTT GAT GGT CTA CAG AAT GAA TGT GCC CTT AGT GCT CTT GGA GAG       1713/524
           W   N   C   D   V   D   G   L   Q   N   E   C   A   L   S   A   L   G   E

1714/525  CTA ATT AAT CAT CTG TCT AGG CTA AAG GAA GAT GTA CTT AAG CAT GGG GAT ATT TTT       1773/544
           L   I   N   H   L   S   R   L   K   E   D   V   L   K   H   G   D   I   F

1774/545  CCA TAC CAA GTT TAC AGG GGT TGT CTC AGA ATT GAT GGC CAG ACG ATG GTA AAT CTT GAG   1833/564
           P   Y   Q   V   Y   R   G   C   L   R   I   D   G   Q   T   M   V   N   L   E
```

Figure 9 (Continued)

```
1834        ATA TTT AAC AAT AGC TGT GAT GGT GGT CCT  TCA GGG ACC TTG TAC AAA TAT CTT GAT AAC        1893
565          I   F   N   N   S   C   D   G   G   P    S   G   T   L   Y   K   Y   L   D   N          584

1894        TGT GTT AGT CCA ACT AAG GGT CGA CTC TTA  AGG AAT TGG ATC TGC CAT CCA CTC AAA GAT        1953
585          C   V   S   P   T   K   G   R   L   L    R   N   W   I   C   H   P   L   K   D          604

1954        GTA GAA AGC ATC AAT AAA CGG CTT GAT GTT  GAA GAA TTC ACG GCA AAC TCA GAA AGT            2013
605          V   E   S   I   N   K   R   L   D   V    E   E   F   T   A   N   S   E   S              624

2014        ATG CAA ATC ACT GGC CAG TAT CTC CAC AAA  CTT CCA GAC TTA GAA AGA CTG CTC GGA CGC        2073
625          M   Q   I   T   G   Q   Y   L   H   K    L   P   D   L   E   R   L   L   G   R          644

2074        ATC AAG TCT AGC GTT CGA TCA GCC TCT GTG  TTG CCT GCT CTT CTG GGG AAA AAA GTG            2133
645          I   K   S   S   V   R   S   A   S   V    L   P   A   L   L   G   K   K   V              664

2134        CTG AAA CAA CGA GTT AAA GCA TTT GGG CAA  ATT GTG AAA GGG TTC AGA AGT GGA ATT GAT        2193
665          L   K   Q   R   V   K   A   F   G   Q    I   V   K   G   F   R   S   G   I   D          684

2194        CTG TTG GCT CTA CAG AAG GAA TCA AAT ATG  ATG AGT TTG CTT TAT AAA CTC TGT AAA            2253
685          L   L   A   L   Q   K   E   S   N   M    M   S   L   L   Y   K   L   C   K              704

2254        CTT CCT ATA TTA GTA GGA AAA AGC GGG CTA  GAG TTA TTT CTT TCT CAA TTC GAA GCC            2313
705          L   P   I   L   V   G   K   S   G   L    E   L   F   L   S   Q   F   E   A              724

2314        ATA GAT AGC GAC TTT CCA AAT TAT CAG AAC  CAA GAT GTG ACA GAA AAC GCT GAA ACT            2373
725          I   D   S   D   F   P   N   Y   Q   N    Q   D   V   T   E   N   A   E   T              744

2374        CTC ACA ATA CTT ATC GAA CTT TTT ATC GAA  AGA TCT CAA TGG TCT GAG GTC ATT CAC            2433
745          L   T   I   L   I   E   L   F   I   E    R   S   Q   W   S   E   V   I   H              764

2434        ACC ATA AGC TGC CTA GAT GTC CTG AGA TCT  TTT GCA ATC GCA AGT CTC TCT GCT GAA ACT        2493
765          T   I   S   C   L   D   V   L   R   S    F   A   I   A   S   L   S   A   E   T          784

2494        AGC ATG GCC AGG CCT GTT ATT TTT CCC GAA  TCA GAA GCT ACA GAT CAG AAT CAG AAA ACA        2553
785          S   M   A   R   P   V   I   F   P   E    S   E   A   T   D   Q   N   Q   K   T          804

2554        AAA GGG CCA ATA CTT AAA ATC CAA GGA CTA  TGG CAT CCA TTT GCA GTT GCA GCC GAT GGT        2613
805          K   G   P   I   L   K   I   Q   G   L    W   H   P   F   A   V   A   A   D   G          824

2614        CAA TTG CCT GTT CCG AAT GAT ATA CTC CTT  GGC GAG GCT AGA AGA AGC AGT AGC ATT            2673
825          Q   L   P   V   P   N   D   I   L   L    G   E   A   R   R   S   S   S   I              844

2674        CAT CCT CGG TCA TTG TTA CTG ACG GGA CCA  AAC ATG GGC GGA AAA TCA ACT CTT CTT CGT        2733
845          H   P   R   S   L   L   L   T   G   P    N   M   G   G   K   S   T   L   L   R          864
```

Figure 9 (Continued)

Figure 9 — nucleotide/amino-acid sequence (continued). Each line spans 20 codons; the numbers at left give the starting nucleotide position (upper) and amino-acid position (lower), and the numbers at right give the ending positions.

```
2734  GCA ACA TGT CTG GCC GTT ATC TTT GCC CAA CTT GGC TGC TAC GTG CCG TGT GAG TCT TGC  2793
 865   A   T   C   L   A   V   I   F   A   Q   L   G   C   Y   V   P   C   E   S   C    884

2794  GAA ATC TCC CTC GTG GAT ACT ATC TTC ACA AGG CTT GGC GCA TCT GAT AGA ATC ATG ACA  2853
 885   E   I   S   L   V   D   T   I   F   T   R   L   G   A   S   D   R   I   M   T    904

2854  GGA GAG AGT ACC TTT TTG GTA GAA TGC CTT CAG AAT GCA ACT AGT ACT TTC GAT GGA TAC  2913
 905   G   E   S   T   F   L   V   E   C   L   Q   N   A   T   S   T   F   D   G   Y    924

2914  CAG GAT TCA CTA GTA ATC CTT GAC CTT GAC GAA CTG CTT GGC AGA GGA ACT TCA ACT TTC  2973
 925   Q   D   S   L   V   I   L   D   L   D   E   L   L   G   R   G   T   S   T   F    944

2974  GCC ATT GCA TAC TCG GTT TTT CGT CAC CTG GTA GAG AAA GTT CAA TGT CGG ATG CTC TTT  3033
 945   A   I   A   Y   S   V   F   R   H   L   V   E   K   V   Q   C   R   M   L   F    964

3034  GCA ACA CAT TAC CAC CCT CTC ACC AAG GAA GAA TTC GCG TCT CAC CCA CGT GTC ACC TCG  3093
 965   A   T   H   Y   H   P   L   T   K   E   E   F   A   S   H   P   R   V   T   S    984

3094  CAC ATG GCT TGC GCA TTC AAA TCA AGA TCT GAT TAT CAA TCT GAT TGT CCT GAG GGA GCT  3153
 985   H   M   A   C   A   F   K   S   R   S   D   Y   Q   S   D   C   P   E   G   A   1004

3154  CTA GTG TTC TTG TAC CGT TTA ACC GAG GGA ACA GCA TCA GGA CTT AGA TCT GAG TTC TCA  3213
1005   L   V   F   L   Y   R   L   T   E   G   T   A   S   G   L   R   S   E   F   S   1024

3214  GCA CTC ATG GCT GGA ATA CCA AAC CAA CAA GTT GGT TGT GAT CAA GAC GTC TCA GCC GCC  3273
1025   A   L   M   A   G   I   P   N   Q   Q   V   G   C   D   Q   D   V   S   A   A   1044

3274  ATG AAG AGA TCA ATT GGG GGA AAC TCA GAG TTC AAG TCA GAG CTA AGA TCT GAG TTC AGT  3333
1045   M   K   R   S   I   G   G   N   S   E   F   K   S   E   L   R   S   E   F   S   1064

3334  CTG CAT GAA GAC TGG CTC AAG TCA TTG GGT ATT TCT CGA GTC GCC CAC AAC AAT GCA GCC  3393
1065   L   H   E   D   W   L   K   S   L   G   I   S   R   V   A   H   N   N   A   A   1084

3394  CCC ATT GGC GAA GAT GAC TAC GAT TTG TTT TGC CTT TGG CAT GAG ATC AAA TCC TCT AGT  3453
1085   P   I   G   E   D   D   Y   D   L   F   C   L   W   H   E   I   K   S   S   S   1104

3454  TAC TGT GTT CCC AAA TAA ATG GCT ATG ACA TAA CACTATCTGAAGCTCGTTAAGTCTTTTGCCTCTCT  3521
1105   Y   C   V   P   K   *   M   A   M   T   *                                          5

3522 G ATG TTT ATT CCT CTT AAA AAA TGC TTA TAT ATC AAA AAA TTG TTT CCT CGA TTA AAA      3579
   1    M   F   I   P   L   K   K   C   L   Y   I   K   K   L   F   P   R   L   K        19

3580  AAA AAA AAA AAA AAA AAA AAA AAA AAA                                                 3606
  20   K   K   K   K   K   K   K   K   K                                                   28
```

Figure 9 (Continued)

Figure 10

```
TTTTTTGGTTGCTAACAATAAAGGTATACGGTTTTATGTCATCAATATAA      50
CTATATATAAAAGAAATGAAAGATATATATTGTTTTTTCATTTATCAAAC      100
AAAACAACAAGACTTTTTTTTTACTTTTTACATTGGTCAACAAAATACAA      150
GATAAACGACATCGTTTAATCATTTCCCAATTTTACCCCTAAGTTTAACA      200
CCTAGAACCTTCTCCATCTTCGCAAGCACAGCCTGATTAGGAACAGCTTT      250
ACCATTCTCATATTCCTGAACTACCTGAGTCCTCTCATTGATCTGTTTCG      300
CCAAATCCGCTTGTGACATCTTCTTCTCCAATCTCGCTTTCTGTATCATC      350
AACCTCACCTCTGCTTTCACACGATCCATCGCCGCAGGCTCTGTTTCTTC      400
TTCCAGCTTCTTCGTGTTAATCACCGGAACCGCCGTAGATTTCCCCTTTT      450
TGTTCGAACCGGCATCGAATTTCTTAACCGTTTGAACCGCGACACCGTTT      500
CTCAGAGCTGCGTTAACCGCTTTCGGATCGCGTAGGTCTTGGCTCTTTTG      550
TTTTGATTTGTGGAGAACTACTGGTTCCCAGTCTTGTGTTACTGCTCCTG      600
GGTATCTGCTCGGCATCGTCGATGAATTGAGAGAAAGGAACAACGCGAAA      650
ATTTTATTAATCTGAGTTTTGAAATTGAGAAACGATGAAGATGAAGAATG      700
TTGTTGAGAGGATTGTGATATTTATATATACGAAGATTGGTTTCTGGAGA      750
ATTCGATCATCTTTTTCTCCATTTTCGTCTCTGGAACGTTCTTAGAGATG      800
ATTGACGACGTGTCATTATCTGATTTGCAGTTAACCAATGCTTTTTGGGT      850
TGGATTCGTGGTACACCATATTATCCGATTTGGCTCAATGGTTTTATATA      900
AATTTGGTTTTCGGTTCGGTTATGAGTTATCATTAAAATTAAGCTAACCA      950
AAAATTTTCGTAAAATTTATTTCGGTTTCAATTCGGATCCCTTACTTCCA      1000
GAACCGAATTATTCGAAACCGGGGTTAGCCGAACCGAATACCAATGCCTG      1050
ATTGACTCGTTGGCTAGAAAGATCCAACGGTATACAATAATAGAACATAA      1100
ATCGGACGGTCATCAAAGCCTCAAAGAGTGAACAGTCAACAAAAAAAGTT      1150
GAGCCCTGAGGAGTATCGTTTCCGCCATTTCTACGACGCAAGGCGAAAAT      1200
TTTTGGCGCCAATCTTTCCCCCCTTTCGAATTCTCTCAGCTCAAAACATC      1250
GTTTCTCTCTCACTCTCTCTCACAATTCCAAAAAATGCAGCGCCAGAGAT      1300
CGATTTTGTCTTTCTTCCAAAAACCCACGGCGGCGACTACGAAGGGTTTG      1350
GTTTCCGGCGATGCTGCTAGCGGCGGGGGCGGCAGCGGAGACCACGATTT      1400
AATGTGAAGGAAGGGGATGCTAAAGGCGACGCTTCTGTACGTTTTGCTGT      1450
TTCGAAATCTGTCGATGAGGTTAGAGGAACGGATACTCCACCGGAGAAGG      1500
TTCCGCGTCGTGTCCTGCCGTCTGGATTTAAGCCGGCTGAATCCGCCGGT      1550
GATGCTTCGTCCCTGTTCTCCAATATTATGCATAAGTTTGTAAAAGTCGA      1600
TGATCGAGATTGTTCTGGAGAGAGGTACTAATCTTCGATTCTCTTAATTT      1650
TGTTATCTTTAGCTGGAAGAAGAAGATTCGTGTAATTTGTTGTATTCGTT      1700
GGAGAGATTCTGATTACTGCATTGGATCGTTGTTTACAAATTTTCAGGAG      1750
CCGAGAAGATGTTGTTCCGCTGAATGATTCATCTCTATGTATGAAGGCTA      1800
ATGATGTTATTCCTCAATTTCGTTCCAATAATGGTAAAACTCAAGAAAGA      1850
AACCATGCTTTTAGTTTCAGTGGGAGAGCTGAACTTAGATCAGTAGAAGA      1900
TATAGGAGTAGATGGCGATGTTCCTGGTCCAGAAACACCAGGGATGCGTC      1950
CACGTGCTTCTCGCTTGAAGCGAGTTCTGGAGGATGAAATGACTTTTAAG      2000
GAGGATAAGGTTCCTGTATTGGACTCTAACAAAAGGCTGAAAATGCTCCA      2050
GGATCCGGTTTGTGGAGAGAAGAAAGAAGTAAACGAAGGAACCAAATTTG      2100
AATGGCTTGAGTCTTCTCGAATCAGGGATGCCAATAGAAGACGTCCTGAT      2150
GATCCCCTTTACGATAGAAAGACCTTACACATACCACCTGATGTTTTTCAA      2200
```

Figure 11

```
GAAAATGTCTGCATCACAAAAGCAATATTGGAGTGTTAAGAGTGAATATA          2250
TGGACATTGTGCTTTTCTTTAAAGTGGTTAGTAACTATTAATCTAGTGTT          2300
CAATCCATTTCCTCAATGTGATTTGTTCACTTACATCTGTTTACGTTATG          2350
CTCTTCTCAGGGGAAATTTTATGAGCTGTATGAGCTAGATGCGGAATTAG          2400
GTCACAAGGAGCTTGACTGGAAGATGACCATGAGTGGTGTGGGAAAATGC          2450
AGACAGGTAAATTAGTTGAAACAACTGGCCTGCTTGAATTATTGTGTCTA          2500
TAAATTTTGACACCACCTTTTGTTTCAGGTTGGTATCTCTGAAAGTGGGA          2550
TAGATGAGGCAGTGCAAAAGCTATTAGCTCGTGGGTAAGGGAACCATCAT          2600
ACTTTATGGAATTCGTTTACTGCTACTTCGGCTAGGATTTAAGAAATGGA          2650
AATCACTTCAAGCATCATTAGTTAGGATCCTGAGAACTCAGGATGTTTTC          2700
TTATTCGTTATATAATAAGTCTTTTCATCAAGGAGTAACAAACAAAACTT          2750
GCACAATATTTGTGTGCTCACTGGCAAGGCATATATACCCAGCTAACCTT          2800
TGCTAGTTCACTGTAGTAACAGTTACGGATAATATATGTTTACTTGTATG          2850
TGGTACCCTCATTTTGTCTCTCATGGAGGCTTTCAAGCCTTGTGTTGAAA          2900
CTGGATAGTTACATATGCTTCCAACAGAAACTAGCATGCAGATTCATATG          2950
CTTTCCTATTCTACTAATTATGTATTGACACACTCGTTGTTTCTTTTGAA          3000
AGATATAAAGTTGGACGAATCGAGCAGCTAGAAACATCTGACCAAGCAAA          3050
AGCCAGAGGTGCTAATACTGTAAGTTTTCTTGGATAGGTCAAGGAGAGTG          3100
TTGCAGACTGTTTTTGATCATTTCTTTTTCTGTACATTACTTTCATGCTG          3150
TAATTAACTCAATGGCTATTCTGGTCTGATTATCAGATAATTCCAAGGAA          3200
GCTAGTTCAGGTATTAACTCCATCAACAGCAAGCGAGGGAAACATCGGGC          3250
CTGATGCCGTCCATCTTCTTGCTATAAAAGAGGTTTGTTATTTACTTATT          3300
TATCTTATCATGTTCAGTTCATCCAAGTCCTGAAAAATTACACTCTTCTT          3350
TACCAATCTTCCATCAAGCTGTGTAAAGGATTTGGAATTAGAAAATCATT          3400
ATTTGATGCTTTGTTTTATATGCAAGAGGTTCCCTTGAAAAGATCTGTTT          3450
AAGATTCTTTGCACTTGAAAAATTCAATCTTTTTAAGTGAATCCCCTACT          3500
TTCTTACAATGATCATAGTCTGCAATTGCATGTCAAGTAATATCATTCCT          3550
TGTTACTGCATCCCCCTCTTTCTTAATGACCATTGTCTATGTTGTGTTTG          3600
TCTCGTGTGCTGGAGAAAATGATAGCTGATCCAAGCTGTACATTATCATG          3650
ATTAAGTAGCTGCTCAGGAATTGCCTTTGGTTACATTGCCTAATGGTTTG          3700
ATGTCAATTTTTCTTCTGAATCTTTATTTTAGATCAAAATGGAGCTACAA          3750
AAGTGTTCAACTGTGTATGGATTTGCTTTTGTTGACTGTGCTGCCTTGAG          3800
GTTTTGGGTTGGGTCCATCAGCGATGATGCATCATGTGCTGCTCTTGGAG          3850
CGTTATTGATGCAGGTAAGCAAGTGTATTCTGTATCTTATGTGTACCATG          3900
TGACTTCCTGTGCATATATTTGGGTTGCAGGAACTAATTCTGAATCACCA          3950
TTTGGTATGTTTTTTCCAGGTTTCTCCAAAGGAAGTGTTATATGACAGTA          4000
AAGGTAAACTGCTTGTATCGCCAGTTGTTTTGTTAAACAGAATTTAAGGT          4050
AAATGACACTGGTTAATTTAAAGTGCATACATGTTGAAATATTGCAGGGC          4100
TATCAAGAGAAGCACAAAAGGCTCTAAGGAAATATACGTTGACAGGTACC          4150
ATTTCAGTAGGCAAGCTAACTGACAATTTAACCGCTCACCGAATGATAGG          4200
TCTCTTAAACATTGCTAATGTAGATGATGTTTATGTTTCAATCTAATAGG          4250
GTCTACGGCGGTACAGTTGGCTCCAGTACCACAAGTAATGGGGGATACAG          4300
ATGCTGCTGGAGTTAGAAATATAATAGAATCTAACGGATACTTTAAAGGT          4350
TCTTCTGAATCATGGAACTGTGCTGTTGATGGTCTAAATGAATGTGATGT          4400
```

Figure 11 (Continued)

```
TGCCCTTAGTGCTCTTGGAGAGCTAATTAATCATCTGTCTAGGCTAAAGG       4450
TGTGTTGGCTTGTTTAGTTTTTGCTTTTCACAAATTAAGCAAAGGAACTT       4500
TTCATAACTTACAGTTTCTATCTACTTGCAGCTAGAAGATGTACTTAAGC       4550
ATGGGGATATTTTTCCATACCAAGTTTACAGGGGTTGTCTCAGAATTGAT       4600
GGCCAGACGATGGTAAATCTTGAGATATTTAACAATAGCTGTGATGGTGG       4650
TCCTTCAGGCAAGTGCATATTTCTTTTTTGATAACTTCAACTAGAGGGCA       4700
GACATAGAAGGAAAAATTCTAATACTTCGTACGGATCTCCAGTAAGTAAT       4750
AGCCGATTTTTGTTTACCTATGTAGGGACCTTGTACAAATATCTTGATAA       4800
CTGTGTTAGTCCAACTGGTAAGCGACTCTTAAGGAATTGGATCTGCCATC       4850
CACTCAAAGATGTAGAAAGCATCAATAAACGGCTTGATGTAGTTGAAGAA       4900
TTCACGGCAAACTCAGAAAGTATGCAAATCACTGGCCAGTATCTCCACAA       4950
ACTTCCAGACTTAGAAAGACTGCTCGGACGCATCAAGTCTAGCGTTCGAT       5000
CATCAGCCTCTGTGTTGCCTGCTCTTCTGGGGAAAAAAGTGCTGAAACAA       5050
CGAGTAAGTATCAATCACAAGTTTTCTGAGTAATGCCTTCCATGAGTAGT       5100
ATAGGACTAAAACATTACGGGTCTAGCTAAAGACTGTTCTCCTTCTTTTG       5150
CAATGTCTGGTTATTCATTACATTTCTCTTAACTTATTGCATTGCAGGTT       5200
AAAGCATTTGGGCAAATTGTGAAAGGGTTCAGAAGTGGAATTGATCTGTT       5250
GTTGGCTCTACAGAAGGAATCAAATATGATGAGTTTGCTTTATAAACTCT       5300
GTAAACTTCCTATATTAGTAGGAAAAAGCGGGCTAGAGTTATTTCTTTCT       5350
CAATTCGAAGCAGCCATAGATAGCGACTTTCCAAATTATCAGGTGCCCAT       5400
CTATCTTTCATACTTTACAACAAAATGTCTGTCACTACTCAAAGCAATGC       5450
ATATGGCTTAGATCTCAACTCACACCCCGAGGATCCTAAAGGGATTTGCT       5500
TTTTATTCCTAATGTTTTTGGATGGTTTGATTTATTTCTAACTTGAACTT       5550
ATTAATCTTGTACCAGAACCAAGATGTGACAGATGAAAACGCTGAAACTC       5600
TCACAATACTTATCGAACTTTTTATCGAAAGAGCAACTCAATGGTCTGAG       5650
GTCATTCACACCATAAGCTGCCTAGATGTCCTGAGATCTTTTGCAATCGC       5700
AGCAAGTCTCTCTGCTGGAAGCATGGCCAGGCCTGTTATTTTTCCCGAAT       5750
CAGAAGCTACAGATCAGAATCAGAAAACAAAAGGGCCAATACTTAAAATC       5800
CAAGGACTATGGCATCCATTTGCAGTTGCAGCCGATGGTCAATTGCCTGT       5850
TCCGAATGATATACTCCTTGGCGAGGCTAGAAGAAGCAGTGGCAGCATTC       5900
ATCCTCGGTCATTGTTACTGACGGGACCAAACATGGGCGGAAAATCAACT       5950
CTTCTTCGTGCAACATGTCTGGCCGTTATCTTTGCCCAAGTTTGTATACT       6000
CGTTAGATAATTACTCTATTCTTTGCAATCAGTTCTTCAACATGAATAAT       6050
AAATTCTGTTTTCTGTCTGCAGCTTGGCTGCTACGTGCCGTGTGAGTCTT       6100
GCGAAATCTCCCTCGTGGATACTATCTTCACAAGGCTTGGCGCATCTGAT       6150
AGAATCATGACAGGAGAGAGTAAGTTTTGTTCTCAAAATACCAATTCCTC       6200
GAACTATTTACTCAGATTTGTCTGATTGGACAAGGTGGTTTTGCTTTTT       6250
TTTAGGTACCTTTTTGGTAGAATGCACTGAGACAGCGTCAGTTCTTCAGA       6300
ATGCAACTCAGGATTCACTAGTAATCCTTGACGAACTGGGCAGAGGAACT       6350
AGTACTTTCGATGGATACGCCATTGCATACTCGGTAACCTGCTCTTCTCC       6400
TTCAACTTATACTTGTTGATCAACAAAAACATGCAATTCATTTTGCTGAA       6450
ACTTATTGATTTATATCAGGTTTTTCGTCACCTGGTAGAGAAAGTTCAAT       6500
GTCGGATGCTCTTTGCAACACATTACCACCCTCTCACCAAGGAATTCGCG       6550
TCTCACCCACGTGTCACCTCGAAACACATGGCTTGCGCATTCAAATCAAG       6600
```

Figure 11 (Continued)

```
ATCTGATTATCAACCACGTGGTTGTGATCAAGACCTAGTGTTCTTGTACC        6650
GTTTAACCGAGGGAGCTTGTCCTGAGAGCTACGGACTTCAAGTGGCACTC        6700
ATGGCTGGAATACCAAACCAAGTGGTTGAAACAGCATCAGGTGCTGCTCA        6750
AGCCATGAAGAGATCAATTGGGGAAAACTTCAAGTCAAGTGAGCTAAGAT        6800
CTGAGTTCTCAAGTCTGCATGAAGACTGGCTCAAGTCATTGGTGGGTATT        6850
TCTCGAGTCGCCCACAACAATGCCCCCATTGGCGAAGATGACTACGACAC        6900
TTTGTTTTGCTTATGGCATGAGATCAAATCCTCTTACTGTGTTCCCAAAT        6950
AAATGGCTATGACATAACACTATCTGAAGCTCGTTAAGTCTTTTGCTTCT        7000
CTGATGTTTATTCCTCTTAAAAAATGCTTATATATCAAAAAATTGTTTCC        7050
TCGATTATAACAAGATTATATATGTATCTGTCGGTTTAGCTATGGTATAT        7100
AATATATGTATGTTCATGAGATTGGTCAAGAGAAATACTCACAAACAGTA        7150
TATTAAGAAGGAAATATGTTTATGCATTAATTTAAGTTTCAAGATAAACT        7200
GCAAATAACCTCGACTAAAGTTGCAAAGACCAAACACAAATTACAAAACT        7250
TATAAGACTTAAGTTCTGAATTCCCTAAAACCAAAAAAAAAAACAGAACA        7300
TATTTGTTGCATCTACAAACAACACAAACCTACATAGTTTATAACTTAC        7350
TCATCACTGAGATTAACATCAGAATCATTCTCCATTTCTTCATCTTCACT        7400
CTCATCATCATCACCACCACCATGATGATTCTCCTCCTCTTCACGTAACC        7450
TAGCAATCTCACTCTGAGCTCTATCAACAATCTGCTTCTTCTGCAACTCC        7500
AAATCTCTCTGAAAATCAGCTCTCATCTTCTCCAACTCCTTCATTTGCTC        7550
TTTCTTACTCTTCTCCATCTTCTCATAAACCTTCCCAAACCTCTCAACAG        7600
AATCCGCCAACATCTTATACGAAGCAGCGTCATTAACCTTCTTCCTCTCG        7650
TACTCAACCTCATCATCCTCATCCTCCTCCTCTTCAGAATCACCAGGACT        7700
ATCCATCATCTCATCAAACCCATTAGACTTATCTAAATAAACCTTAGTGT        7750
TCATAAACACAAACTCACCTGAATCAACACCACAAGCTAAACCTAAATCC        7800
GACTTGGGCGAAACACAAAGCAACATATCCAACTTATTGAAAAACGACCA        7850
TTTACTTGAACCTAAACCTGATTTCTCAACCTTAATCTTCTCTTTTCTAT        7900
ACTTCCTCTTCAAGTCATCAATCATTCTCCTACATTGCGTCTCAGATTTC        7950
TCCATCCTTAGCTCCTCACTCACTTTCTCAGCTACTTCATTCCAATCCTC        8000
GTTCCTCAAACTCCTTCTACCCAATTGCAAAAACCTATCTCCCCAAACTT        8050
CAAGCAACACAA                                             8062
```

Figure 11 (Continued)
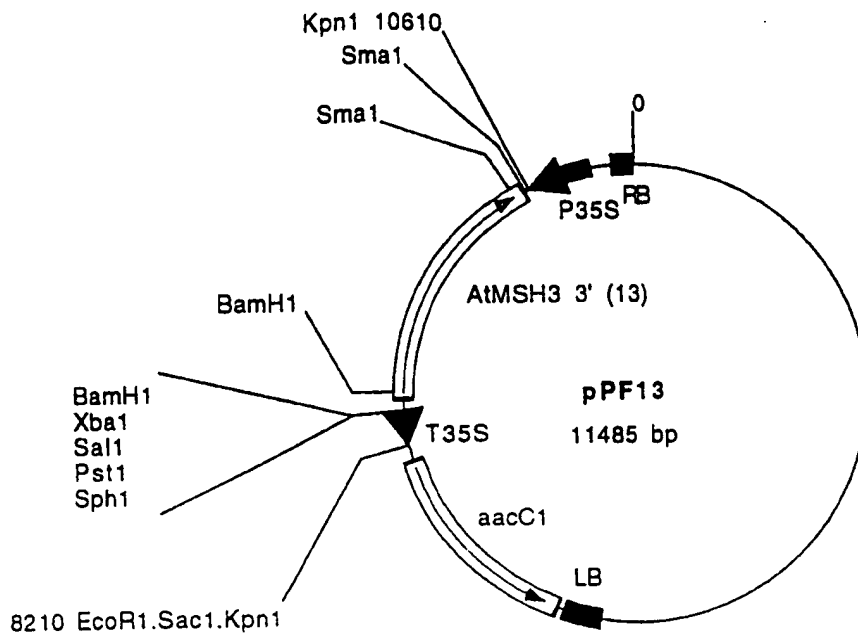
Figure 12

Comments/References: AtMSH3 3' side antisense : AtMSH3 3' (13 = 2104bp) from pUC18/13 Sal1/Sst1/T4 into pCW164 BamH1/T4   in Agrobacterium LBA4404
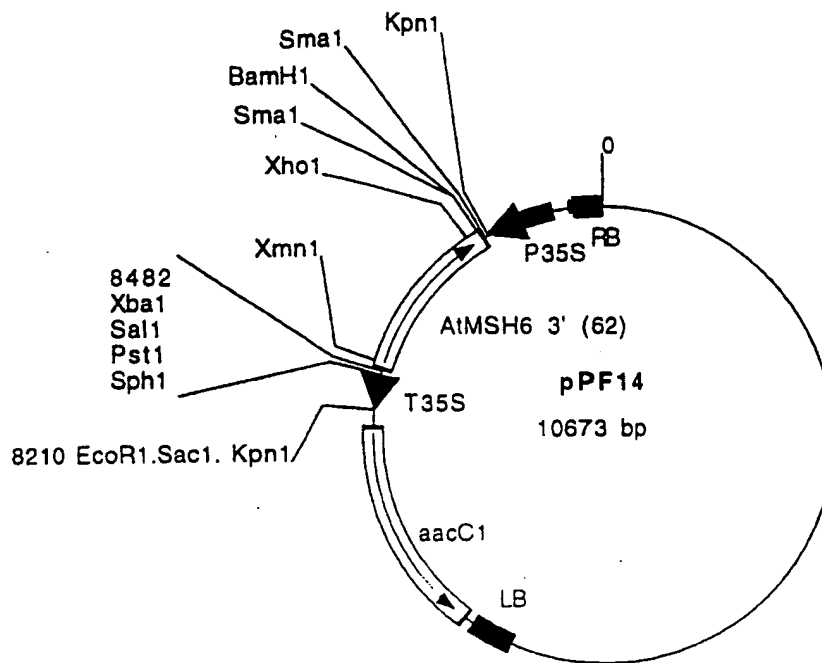
Figure 13

Comments/References: AtMSH6 (S8) 3' side antisens : 62 Sal1/Sst1/T4 (1379bp) into pCW164 BamH1/T4
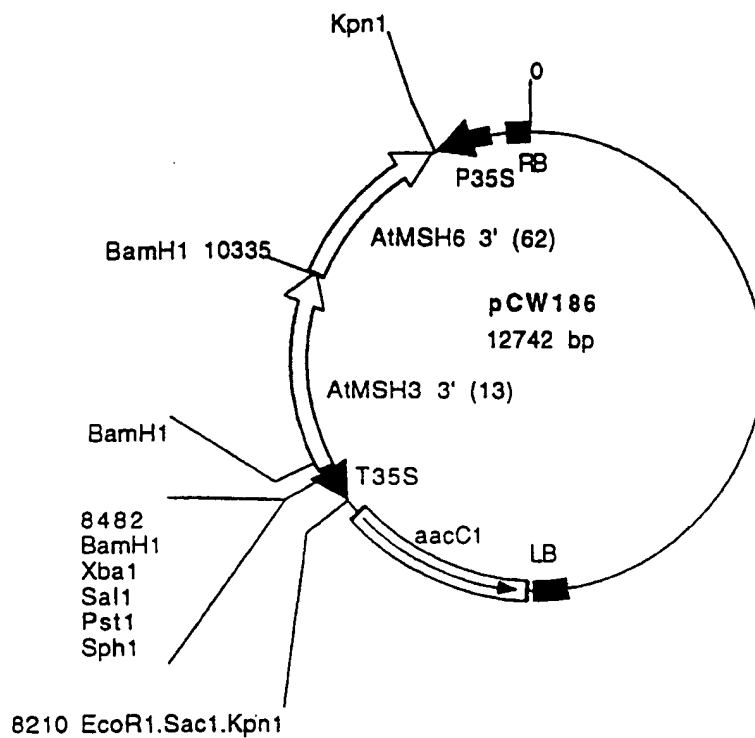
Figure 14

Comments/References: AtMSH6 3'/AtMSH3 3' antisense : AtMSH6 (S8) 3' side (62=1379bp) Sal1/Sst1/T4 into pPF13 (pCW164 AtMSH3 (S5) 3' side (13=2104) antisens)/Sma1. in LBA4404
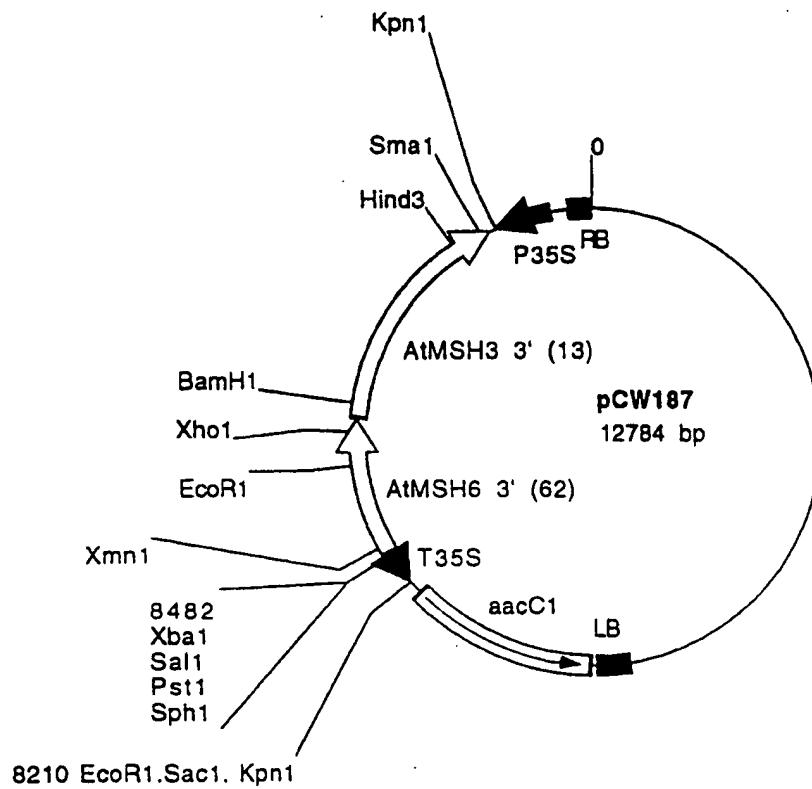
Figure 15

Comments/References: AtMSH3 3'/AtMSH6 3' antisens (D) : AtMSH3 (S5) 3' side (13=2104bp) Sal1/Sst1/T4 into pPF14 (AtMSH6 (S8) 3'side (62=1379bp) antisense into pCW164)/Sma1. in LBA4404
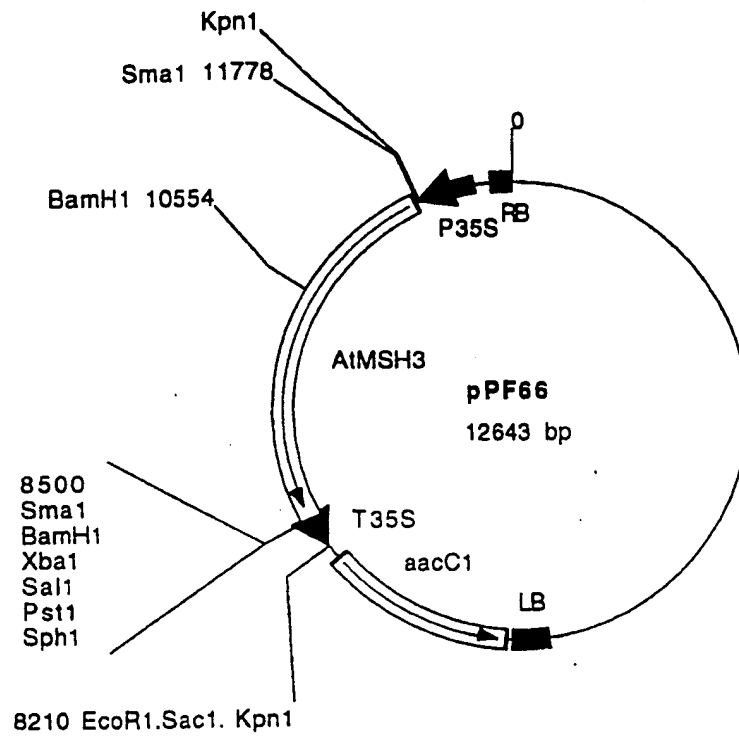
Figure 16

Comments/References: AtMSH3 (S8) complete, sense orientation : pPF26 (3342bp)
Sma1 into pCW164 Sma1
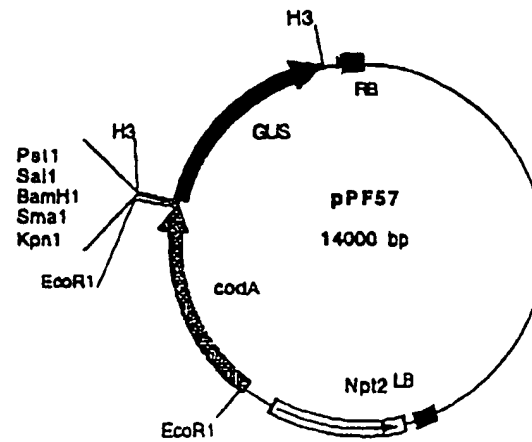
Figure 17

Comments/References: pPZP111 with codA EcoR1 cassette in EcoR1 site and HInd3 GUS cassette in HInd3 site. KanR. All genes under Promoter/terminator 35S

Figure 18
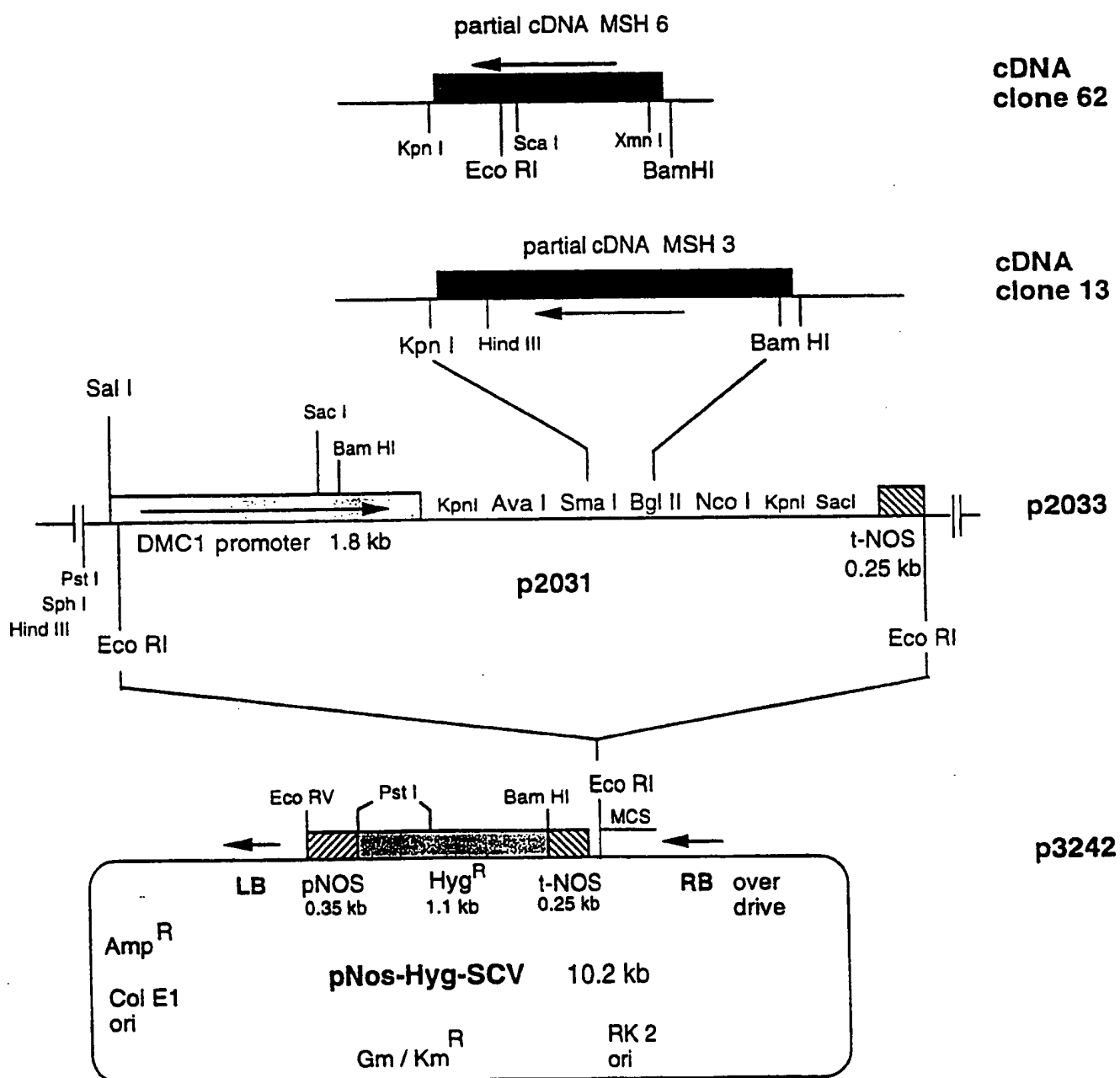
Figure 19

p3243
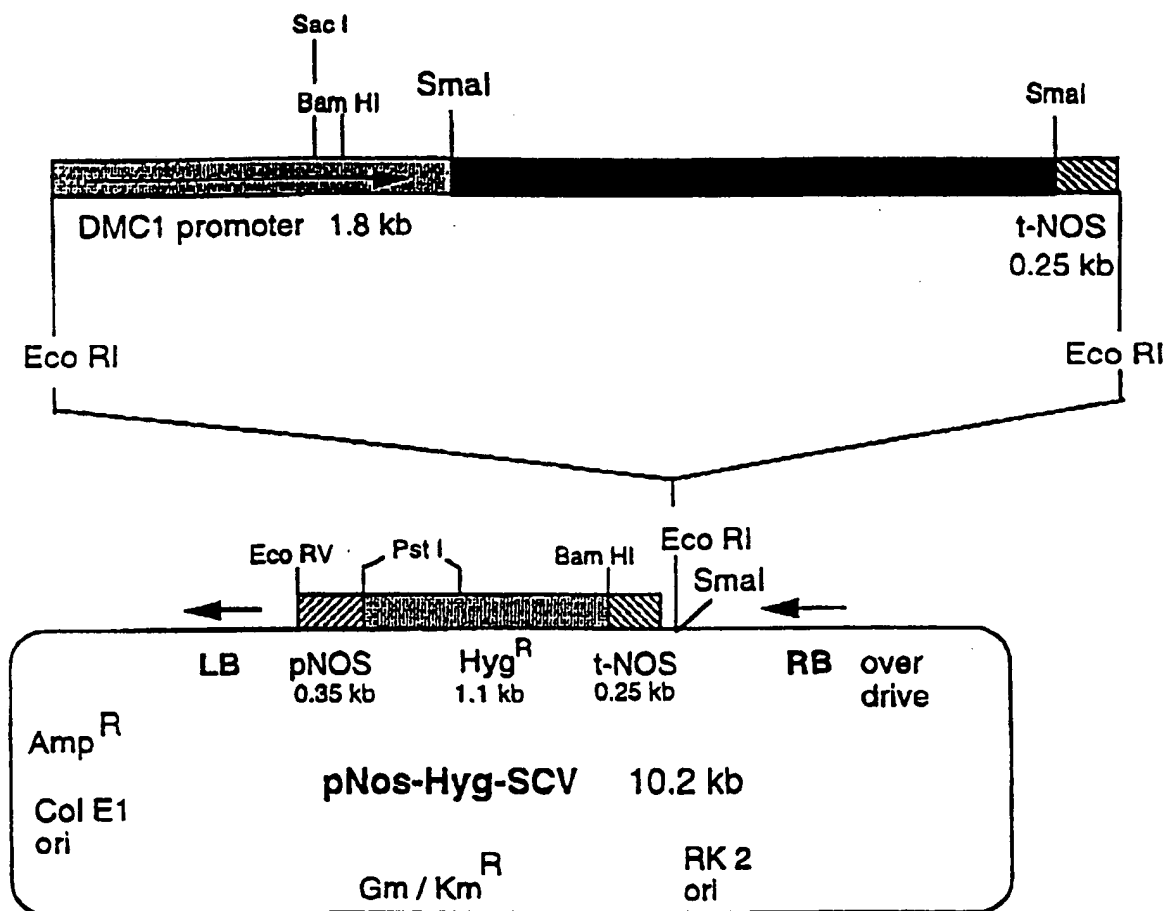
SEQUENCE LISTING

<110>   Rhone-Poulenc Agro; Betzner, Andreas Stefan; Doutriaux,
             Marie-Pascale; Freyssinet, Georges; Perez, Pascual.

<120>   Methods for obtaining plant varieties

<130>   395498C

<150>   PO9745
<151>   1997-10-10

<160>   98

<210>   1
<211>   23
<212>   DNA
<213>   Artificial sequence

<220>
<221>   modified_base
<222>   11
<223>   I

<220>
<221>   modified_base
<222>   14
<223>   I

<220>
<221>   modified_base
<222>   17
<223>   I

<220>
<223>   Degenerate oligonucleotides UPMU used to isolate AtMSH3 and
             AtMSH6.

<300>
<301>   Reenan and Kolodner
<302>   Genetics
<303>   132
<306>   963-973
<307>   1992

<400>   1

ctggatccac nggnccnaay atg   23

<210>   2
<211>   23
<212>   DNA

```
<213>        Artificial sequence

<220>
<221>        modified_base
<222>        15
<223>        I

<220>
<221>        modified_base
<222>        18
<223>        I

<220>
<223>        Degenerate oligonucleotides DOMU used to isolate AtMSH3 and
             AtMSH6.

<300>
<301>        Reenan and Kolodner
<302>        Genetics
<303>        132
<306>        963-973
<307>        1992

<400>        2

ctggatccrt artgngtnrc raa                                           23


<210>        3
<211>        24
<212>        DNA
<213>        Artificial sequence

<220>
<223>        MSH3 specific primer 636 for PCR using cDNA of Arabidopsis
             thaliana ecotype Columbia

<400>        3

tgctagtgcc tcttgcaagc tcat                                         24


<210>        4
<211>        27
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Primer AP1 for PCR using cDNA of Arabidopsis thaliana ecotype
             Columbia containing adapter sequences ligated to both its
             ends

<400>        4
```

ccatcctaat acgactcact atagggc   27

<210>	5
<211>	23
<212>	DNA
<213>	Artificial sequence

<220>
<223>   Primer AP2 for PCR using cDNA of Arabidopsis thaliana ecotype
           Columbia containing adapter sequences ligated to both its
           ends

<400>   5

actcactata gggctcgagc ggc   23

<210>	6
<211>	30
<212>	DNA
<213>	Artificial sequence

<220>
<223>   MSH3 specific primer S525 for PCR using cDNA of Arabidopsis
           thaliana ecotype Columbia

<400>   6

aggttctgat tatgtgtgac gctttactta   30

<210>	7
<211>	29
<212>	DNA
<213>	Artificial sequence

<220>
<223>   MSH3 specific primer S51 for PCR using cDNA of Arabidopsis
           thaliana ecotype Columbia

<400>   7

ggatcgggta ctgggttttg agtgtgagg   29

<210>	8
<211>	24
<212>	DNA
<213>	Artificial sequence

<220>
<223>   MSH3 specific primer 635 for PCR using cDNA of Arabidopsis
           thaliana ecotype Columbia

4

```
<400>            8


gcacgtgctt gatggtgttt tcac                                               24



<210>            9
<211>            28
<212>            DNA
<213>            Artificial sequence

<220>
<223>            MSH3 specific primer S523 for PCR using cDNA of Arabidopsis
                 thaliana ecotype Columbia

<400>            9


tcagacagta tccagcatgg cagaagta                                          28



<210>            10
<211>            33
<212>            DNA
<213>            Artificial sequence

<220>
<223>            MSH3 specific primer 1S5 for PCR using cDNA of Arabidopsis
                 thaliana ecotype Columbia

<400>            10


atcccgggat gggcaagcaa aagcagcaga cga                                    33



<210>            11
<211>            27
<212>            DNA
<213>            Artificial sequence

<220>
<223>            MSH3 specific primer S53 for PCR using cDNA of Arabidopsis
                 thaliana ecotype Columbia

<400>            11


gacaaagagc gaaatgaggc cccttgg                                           27



<210>            12
<211>            1250
<212>            DNA
<213>            Arabidopsis thaliana ecotype Columbia
<223>            Clone 52
```

<400>   12

cccgggatgg gcaagcaaaa gcagcagacg atttctcgtt tcttcgctcc caaacccaaa   60

tccccgactc acgaaccgaa tccggtagcc gaatcatcaa caccgccacc gaagatatcc   120

gccactgtat ccttctctcc ttccaagcgt aagcttctct ccgaccacct cgccgccgcg   180

tcacccaaaa agcctaaact ttctcctcac actcaaaacc cagtacccga tcccaattta   240

caccaaagat ttctccagag atttctggaa ccctcgccgg aggaatatgt tcccgaaacg   300

tcatcatcga ggaaatacac accattggaa cagcaagtgg tggagctaaa gagcaagtac   360

ccagatgtgg ttttgatggt ggaagttggt tacaggtaca gattcttcgg agaagacgcg   420

gagatcgcag cacgcgtgtt gggtatttac gctcatatgg atcacaattt catgacggcg   480

agtgtgccaa catttcgatt gaatttccat gtgagaagac tggtgaatgc aggatacaag   540

attggtgtag tgaagcagac tgaaactgca gccattaagt cccatggtgc aaaccggacc   600

ggcccttttt tccggggact gtcggcgttg tataccaaag ccacgcttga agcggctgag   660

gatataagtg gtggttgtgg tggtgaagaa ggttttggtt cacagagtaa tttcttggtt   720

tgtgttgtgg atgagagagt taagtcggag acattaggct gtggtattga aatgagtttt   780

gatgttagag tcggtgttgt tggcgttgaa atttcgacag gtgaagttgt ttatgaagag   840

ttcaatgata atttcatgag aagtggatta gaggctgtga ttttgagctt gtcaccagct   900

gagctgttgc ttggccagcc tctttcacaa caaactgaga gtttttggt ggcacatgct   960

ggacctacct caaacgttcg agtggaacgt gcctcactgg attgtttcag caatggtaat   1020

gcagtagatg aggttatttc attatgtgaa aaaatcagcg caggtaactt agaagatgat   1080

aaagaaatga agctggaggc tgctgaaaaa ggaatgtctt gcttgacagt tcatacaatt   1140

atgaacatgc cacatctgac tgttcaagcc ctcgccctaa cgttttgcca tctcaaacag   1200

tttggatttg aaaggatcct ttaccaaggg gcctcatttc gctctttgtc   1250

<210>   13
<211>   34
<212>   DNA
<213>   Artificial sequence

<220>
<223>   MSH3 specific primer 2S5 for PCR using cDNA of Arabidopsis
            thaliana ecotype Columbia

<400>   13

atcccgggtc aaaatgaaca agttggtttt agtc   34

<210>   14
<211>   27
<212>   DNA
<213>   Artificial sequence

<220>
<223>   MSH3 specific primer S52 for PCR using cDNA of Arabidopsis
            thaliana ecotype Columbia

<400>   14

gccacatctg actgttcaag ccctcgc   27

<210>   15
<211>   2110
<212>   DNA
<213>   *Arabidopsis thaliana* ecotype Columbia
<223>   Clone 13

<400>   15

gccacatctg actgttcaag ccctcgccct aacgttttgc catctcaaac agtttggatt   60

tgaaaggatc ctttaccaag gggcctcatt tcgctctttg tcaagtaaca cagagatgac   120

tctctcagcc aatactctgc aacagttgga ggttgtgaaa ataattcag atggatcgga   180

atctggctcc ttattccata atatgaatca cacacttaca gtatatggtt ccaggcttct   240

tagacactgg gtgactcatc ctctatgcga tagaaatttg atatctgctc ggcttgatgc   300

tgtttctgag atttctgctt gcatgggatc tcatagttct tcccagctca gcagtgagtt   360

ggttgaagaa ggttctgaga gagcaattgt atcacctgag ttttatctcg tgctctcctc   420

agtcttgaca gctatgtcta gatcatctga tattcaacgt ggaataacaa gaatctttca   480

tcggactgct aaagccacag agttcattgc agttatggaa gctatttta ttgcggggaa   540

gcaaattcag cggcttggca taaagcaaga ctctgaaatg aggagtatgc aatctgcaac   600

tgtgcgatct actcttttga gaaaattgat ttctgttatt tcatcccctg ttgtggttga   660

caatgccgga aaacttctct ctgccctaaa taggaagcg ctgttcgag gtgacttgct   720

cgacatacta atcacttcca gcgaccaatt tcctgagctt gctgaagctc gccaagcagt   780

tttagtcatc agggaaaagc tggattcctc gatagcttca tttcgcaaga agctcgctat   840

```
tcgaaatttg gaatttcttc aagtgtcggg gatcacacat ttgatagagc tgcccgttga     900

ttccaaggtc cctatgaatt gggtgaaagt aaatagcacc aagaagacta ttcgatatca     960

tcccccagaa atagtagctg gcttggatga gctagctcta gcaactgaac atcttgccat    1020

tgtgaaccga gcttcgtggg atagtttcct caagagtttc agtagatact acacagattt    1080

taaggctgcc gttcaagctc ttgctgcact ggactgtttg cactcccttt caactctatc    1140

tagaaacaag aactatgtcc gtcccgagtt tgtggatgac tgtgaaccag ttgagataaa    1200

catacagtct ggtcgtcatc ctgtactgga gactatatta caagataact tcgtcccaaa    1260

tgacacaatt ttgcatgcag aaggggaata ttgccaaatt atcaccggac ctaacatggg    1320

aggaaagagc tgctatatcc gtcaagttgc tttaatttcc ataatggctc aggttggttc    1380

ctttgtacca gcgtcattcg ccaagctgca cgtgcttgat ggtgttttca ctcggatggg    1440

tgcttcagac agtatccagc atggcagaag tacctttcta gaagaattaa gtgaagcgtc    1500

acacataatc agaacctgtt cttctcgttc gcttgttata ttagatgagc ttggaagagg    1560

cactagcaca cacgacggtg tagccattgc ctatgcaaca ttacagcatc tcctagcaga    1620

aaagagatgt ttggttcttt ttgtcacgca ttaccctgaa atagctgaga tcagtaacgg    1680

attcccaggt tctgttggga cataccatgt ctcgtatctg acattgcaga aggataaagg    1740

cagttatgat catgatgatg tgacctacct atataagctt gtgcgtggtc tttgcagcag    1800

gagctttggt tttaaggttg ctcagcttgc ccagatacct ccatcatgta tacgtcgagc    1860

catttcaatg gctgcaaaat tggaagctga ggtacgtgca agagagagaa atacacgcat    1920

gggagaacca gaaggacatg aagaaccgag aggcgcagaa gaatctattt cggctctagg    1980

tgacttgttt gcagacctga aatttgctct ctctgaagag gacccttgga aagcattcga    2040

gttttttaaag catgcttgga agattgctgg caaaatcaga ctaaaaccaa cttgttcatt    2100

ttgacccggg                                                           2110
```

```
<210>       16
<211>       29
<212>       DNA
<213>       Artificial sequence

<220>
<223>       MSH3 specific primer S51 for PCR using cDNA of Arabidopsis
            thaliana ecotype Columbia

<400>       16
```

ggatcgggta ctgggttttg agtgtgagg   29

<210>   17
<211>   30
<212>   DNA
<213>   Artificial sequence

<220>
<223>   MSH3 specific primer S525 for PCR using cDNA of Arabidopsis
             thaliana ecotype Columbia

<400>   17

aggttctgat tatgtgtgac gctttactta   30

<210>   18
<211>   3522
<212>   DNA
<213>   Arabidopsis thaliana ecotype Columbia

<220>
<221>   CDS
<222>   (100)....(3342)
<223>   AtMSH3 full-length cDNA and deduced sequence of the encoded
             polypeptide

<400>   18

cctaagaaag cgcgcgaaaa ttggcaaccc aagttcgcca tagccacgac cacgaccttc   60

catttctctt aaacggagga gattacgaat aaagcaatt   99

atg ggc aag caa aag cag cag acg att tct cgt ttc ttc gct ccc aaa   147
Met Gly Lys Gln Lys Gln Gln Thr Ile Ser Arg Phe Phe Ala Pro Lys
 1   5   10   15

ccc aaa tcc ccg act cac gaa ccg aat ccg gta gcc gaa tca tca aca   195
Pro Lys Ser Pro Thr His Glu Pro Asn Pro Val Ala Glu Ser Ser Thr
            20   25   30

ccg cca ccg aag ata tcc gcc act gta tcc ttc tct cct tcc aag cgt   243
Pro Pro Pro Lys Ile Ser Ala Thr Val Ser Phe Ser Pro Ser Lys Arg
        35   40   45

aag ctt ctc tcc gac cac ctc gcc gcc gcg tca ccc aaa aag cct aaa   291
Lys Leu Leu Ser Asp His Leu Ala Ala Ala Ser Pro Lys Lys Pro Lys
    50   55   60

ctt tct cct cac act caa aac cca gta ccc gat ccc aat tta cac caa   339
Leu Ser Pro His Thr Gln Asn Pro Val Pro Asp Pro Asn Leu His Gln
65   70   75   80

9

```
aga ttt ctc cag aga ttt ctg gaa ccc tcg ccg gag gaa tat gtt ccc      387
Arg Phe Leu Gln Arg Phe Leu Glu Pro Ser Pro Glu Glu Tyr Val Pro
            85              90                  95

gaa acg tca tca tcg agg aaa tac aca cca ttg gaa cag caa gtg gtg      435
Glu Thr Ser Ser Ser Arg Lys Tyr Thr Pro Leu Glu Gln Gln Val Val
            100             105                 110

gag cta aag agc aag tac cca gat gtg gtt ttg atg gtg gaa gtt ggt      483
Glu Leu Lys Ser Lys Tyr Pro Asp Val Val Leu Met Val Glu Val Gly
            115             120                 125

tac agg tac aga ttc ttc gga gaa gac gcg gag atc gca gca cgc gtg      531
Tyr Arg Tyr Arg Phe Phe Gly Glu Asp Ala Glu Ile Ala Ala Arg Val
    130             135                 140

ttg ggt att tac gct cat atg gat cac aat ttc atg acg gcg agt gtg      579
Leu Gly Ile Tyr Ala His Met Asp His Asn Phe Met Thr Ala Ser Val
145             150                 155                 160

cca aca ttt cga ttg aat ttc cat gtg aga aga ctg gtg aat gca gga      627
Pro Thr Phe Arg Leu Asn Phe His Val Arg Arg Leu Val Asn Ala Gly
                165             170                 175

tac aag att ggt gta gtg aag cag act gaa act gca gcc att aag tcc      675
Tyr Lys Ile Gly Val Val Lys Gln Thr Glu Thr Ala Ala Ile Lys Ser
                180             185                 190

cat ggt gca aac cgg acc ggc cct ttt ttc cgg gga ctg tcg gcg ttg      723
His Gly Ala Asn Arg Thr Gly Pro Phe Phe Arg Gly Leu Ser Ala Leu
            195             200                 205

tat acc aaa gcc acg ctt gaa gcg gct gag gat ata agt ggt ggt tgt      771
Tyr Thr Lys Ala Thr Leu Glu Ala Ala Glu Asp Ile Ser Gly Gly Cys
    210             215                 220

ggt ggt gaa gaa ggt ttt ggt tca cag agt aat ttc ttg gtt tgt gtt      819
Gly Gly Glu Glu Gly Phe Gly Ser Gln Ser Asn Phe Leu Val Cys Val
225             230                 235                 240

gtg gat gag aga gtt aag tcg gag aca tta ggc tgt ggt att gaa atg      867
Val Asp Glu Arg Val Lys Ser Glu Thr Leu Gly Cys Gly Ile Glu Met
                245             250                 255

agt ttt gat gtt aga gtc ggt gtt gtt ggc gtt gaa att tcg aca ggt      915
Ser Phe Asp Val Arg Val Gly Val Val Gly Val Glu Ile Ser Thr Gly
            260             265                 270

gaa gtt gtt tat gaa gag ttc aat gat aat ttc atg aga agt gga tta      963
Glu Val Val Tyr Glu Glu Phe Asn Asp Asn Phe Met Arg Ser Gly Leu
            275             280                 285
```

10

```
gag gct gtg att ttg agc ttg tca cca gct gag ctg ttg ctt ggc cag      1011
Glu Ala Val Ile Leu Ser Leu Ser Pro Ala Glu Leu Leu Leu Gly Gln
    290             295             300

cct ctt tca caa caa act gag aag ttt ttg gtg gca cat gct gga cct      1059
Pro Leu Ser Gln Gln Thr Glu Lys Phe Leu Val Ala Met Ala Gly Pro
305             310             315             320

acc tca aac gtt cga gtg gaa cgt gcc tca ctg gat tgt ttc agc aat      1107
Thr Ser Asn Val Arg Val Glu Arg Ala Ser Leu Asp Cys Phe Ser Asn
                325             330             335

ggt aat gca gta gat gag gtt att tca tta tgt gaa aaa atc agc gca      1155
Gly Asn Ala Val Asp Glu Val Ile Ser Leu Cys Glu Lys Ile Ser Ala
                340             345             350

ggt aac tta gaa gat gat aaa gaa atg aag ctg gag gct gct gaa aaa      1203
Gly Asn Leu Glu Asp Asp Lys Glu Met Lys Leu Glu Ala Ala Glu Lys
                355             360             365

gga atg tct tgc ttg aca gtt cat aca att atg aac atg cca cat ctg      1251
Gly Met Ser Cys Leu Thr Val His Thr Ile Met Asn Met Pro His Leu
    370             375             380

act gtt caa gcc ctc gcc cta acg ttt tgc cat ctc aaa cag ttt gga      1299
Thr Val Gln Ala Leu Ala Leu Thr Phe Cys His Leu Lys Gln Phe Gly
385             390             395             400

ttt gaa agg atc ctt tac caa ggg gcc tca ttt cgc tct ttg tca agt      1347
Phe Glu Arg Ile Leu Tyr Gln Gly Ala Ser Phe Arg Ser Leu Ser Ser
                405             410             415

aac aca gag atg act ctc tca gcc aat act ctg caa cag ttg gag gtt      1395
Asn Thr Glu Met Thr Leu Ser Ala Asn Thr Leu Gln Gln Leu Glu Val
                420             425             430

gtg aaa aat aat tca gat gga tcg gaa tct ggc tcc tta ttc cat aat      1443
Val Lys Asn Asn Ser Asp Gly Ser Glu Ser Gly Ser Leu Phe His Asn
                435             440             445

atg aat cac aca ctt aca gta tat gct tcc agg ctt ctt aga cac tgg      1491
Met Asn His Thr Leu Thr Val Tyr Gly Ser Arg Leu Leu Arg His Trp
    450             455             460

gtg act cat cct cta tgc gat aga aat ttg ata tct gct cgg ctt gat      1539
Val Thr His Pro Leu Cys Asp Arg Asn Leu Ile Ser Ala Arg Leu Asp
465             470             475             480

gct gtt tct gag att tct gct tgc atg gga tct cat agt tct tcc cag      1587
Ala Val Ser Glu Ile Ser Ala Cys Met Gly Ser His Ser Ser Ser Gln
                485             490             495
```

11

```
ctc agc agt gag ttg gtt gaa gaa ggt tct gag aga gca att gta tca     1635
Leu Ser Ser Glu Leu Val Glu Glu Gly Ser Glu Arg Ala Ile Val Ser
        500                 505                 510

cct gag ttt tat ctc gtg ctc tcc tca gtc ttg aca gct atg tct aga     1683
Pro Glu Phe Tyr Leu Val Leu Ser Ser Val Leu Thr Ala Met Ser Arg
        515                 520                 525

tca tct gat att caa cgt gga ata aca aga atc ttt cat cgg act gct     1731
Ser Ser Asp Ile Gln Arg Gly Ile Thr Arg Ile Phe His Arg Thr Ala
        530                 535                 540

aaa gcc aca gag ttc att gca gtt atg gaa gct att tta ctt gcg ggg     1779
Lys Ala Thr Glu Phe Ile Ala Val Met Glu Ala Ile Leu Leu Ala Gly
545                 550                 555                 560

aag caa att cag cgg ctt ggc ata aag caa gac tct gaa atg agg agt     1827
Lys Gln Ile Gln Arg Leu Gly Ile Lys Gln Asp Ser Glu Met Arg Ser
            565                 570                 575

atg caa tct gca act gtg cga tct act ctt ttg aga aaa ttg att tct     1875
Met Gln Ser Ala Thr Val Arg Ser Thr Leu Leu Arg Lys Leu Ile Ser
            580                 585                 590

gtt att tca tcc cct gtt gtg gtt gac aat gcc gga aaa ctt ctc tct     1923
Val Ile Ser Ser Pro Val Val Val Asp Asn Ala Gly Lys Leu Leu Ser
            595                 600                 605

gcc cta aat aag gaa gcg gct gtt cga ggt gac ttg ctc gac ata cta     1971
Ala Leu Asn Lys Glu Ala Ala Val Arg Gly Asp Leu Leu Asp Ile Leu
        610                 615                 620

atc act tcc agc gac caa ttt cct gag ctt gct gaa gct cgc caa gca     2019
Ile Thr Ser Ser Asp Gln Phe Pro Glu Leu Ala Glu Ala Arg Gln Ala
625                 630                 635                 640

gtt tta gtc atc agg gaa aag ctg gat tcc tcg ata gct tca ttt cgc     2067
Val Leu Val Ile Arg Glu Lys Leu Asp Ser Ser Ile Ala Ser Phe Arg
            645                 650                 655

aag aag ctc gct att cga aat ttg gaa ttt ctt caa gtg tcg ggg atc     2115
Lys Lys Leu Ala Ile Arg Asn Leu Glu Phe Leu Gln Val Ser Gly Ile
            660                 665                 670

aca cat ttg ata gag ctg ccc gtt gat tcc aag gtc cct atg aat tgg     2163
Thr His Leu Ile Glu Leu Pro Val Asp Ser Lys Val Pro Met Asn Trp
        675                 680                 685

gtg aaa gta aat agc acc aag aag act att cga tat cat ccc cca gaa     2211
Val Lys Val Asn Ser Thr Lys Lys Thr Ile Arg Tyr His Pro Pro Glu
        690                 695                 700
```

12

```
ata gta gct ggc ttg gat gag cta gct cta gca act gaa cat ctt gcc        2259
Ile Val Ala Gly Leu Asp Glu Leu Ala Leu Ala Thr Glu His Leu Ala
705             710             715             720

att gtg aac cga gct tcg tgg gat agt ttc ctc aag agt ttc agt aga        2307
Ile Val Asn Arg Ala Ser Trp Asp Ser Phe Leu Lys Ser Phe Ser Arg
                725             730             735

tac tac aca gat ttt aag gct gcc gtt caa gct ctt gct gca ctg gac        2355
Tyr Tyr Thr Asp Phe Lys Ala Ala Val Gln Ala Leu Ala Ala Leu Asp
            740             745             750

tgt ttg cac tcc ctt tca act cta tct aga aac aag aac tat gtc cgt        2403
Cys Leu His Ser Leu Ser Thr Leu Ser Arg Asn Lys Asn Tyr Val Arg
            755             760             765

ccc gag ttt gtg gat gac tgt gaa cca gtt gag ata aac ata cag tct        2451
Pro Glu Phe Val Asp Asp Cys Glu Pro Val Glu Ile Asn Ile Gln Ser
            770             775             780

ggt cgt cat cct gta ctg gag act ata tta caa gat aac ttc gtc cca        2499
Gly Arg His Pro Val Leu Glu Thr Ile Leu Gln Asp Asn Phe Val Pro
785             790             795             800

aat gac aca att ttg cat gca gaa ggg gaa tat tgc caa att atc acc        2547
Asn Asp Thr Ile Leu His Ala Glu Gly Glu Tyr Cys Gln Ile Ile Thr
                805             810             815

gga cct aac atg gga gga aag agc tgc tat atc cgt caa gtt gct tta        2595
Gly Pro Asn Met Gly Gly Lys Ser Cys Tyr Ile Arg Gln Val Ala Leu
                820             825             830

att tcc ata atg gct cag gtt ggt tcc ttt gta cca gcg tca ttc gcc        2643
Ile Ser Ile Met Ala Gln Val Gly Ser Phe Val Pro Ala Ser Phe Ala
                835             840             845

aag ctg cac gtg ctt gat ggt gtt ttc act cgg atg ggt gct tca gac        2691
Lys Leu His Val Leu Asp Gly Val Phe Thr Arg Met Gly Ala Ser Asp
            850             855             860

agt atc cag cat ggc aga agt acc ttt cta gaa gaa tta agt gaa gcg        2739
Ser Ile Gln His Gly Arg Ser Thr Phe Leu Glu Glu Leu Ser Glu Ala
865             870             875             880

tca cac ata atc aga acc tgt tct tct cgt tcg ctt gtt ata tta gat        2787
Ser His Ile Ile Arg Thr Cys Ser Ser Arg Ser Leu Val Ile Leu Asp
                885             890             895

gag ctt gga aga ggc act agc aca cac gac ggt gta gcc att gcc tat        2835
Glu Leu Gly Arg Gly Thr Ser Thr His Asp Gly Val Ala Ile Ala Tyr
            900             905             910
```

```
gca aca tta cag cat ctc cta gca gaa aag aga tgt ttg gtt ctt ttt    2883
Ala Thr Leu Gln His Leu Leu Ala Glu Lys Arg Cys Leu Val Leu Phe
        915              920              925

gtc acg cat tac cct gaa ata gct gag atc agt aac gga ttc cca ggt    2931
Val Thr His Tyr Pro Glu Ile Ala Glu Ile Ser Asn Gly Phe Pro Gly
        930              935              940

tct gtt ggg aca tac cat gtc tcg tat ctg aca ttg cag aag gat aaa    2979
Ser Val Gly Thr Tyr His Val Ser Tyr Leu Thr Leu Gln Lys Asp Lys
945              950              955              960

ggc agt tat gat cat gat gat gtg acc tac cta tat aag ctt gtg cgt    3027
Gly Ser Tyr Asp His Asp Asp Val Thr Tyr Leu Tyr Lys Leu Val Arg
                965              970              975

ggt ctt tgc agc agg agc ttt ggt ttt aag gtt gct cag ctt gcc cag    3075
Gly Leu Cys Ser Arg Ser Phe Gly Phe Lys Val Ala Gln Leu Ala Gln
        980              985              990

ata cct cca tca tgt ata cgt cga gcc att tca atg gct gca aaa ttg    3123
Ile Pro Pro Ser Cys Ile Arg Arg Ala Ile Ser Met Ala Ala Lys Leu
        995              1000             1005

gaa gct gag gta cgt gca aga gag aga aat aca cgc atg gga gaa cca    3171
Glu Ala Glu Val Arg Ala Arg Glu Arg Asn Thr Arg Met Gly Glu Pro
    1010             1015             1020

gaa gga cat gaa gaa ccg aga ggc gca gaa gaa tct att tcg gct cta    3219
Glu Gly His Glu Glu Pro Arg Gly Ala Glu Glu Ser Ile Ser Ala Leu
1025             1030             1035             1040

ggt gac ttg ttt gca gac ctg aaa ttt gct ctc tct gaa gag gac cct    3267
Gly Asp Leu Phe Ala Asp Leu Lys Phe Ala Leu Ser Glu Glu Asp Pro
            1045             1050             1055

tgg aaa gca ttc gag ttt tta aag cat gct tgg aag att gct ggc aaa    3315
Trp Lys Ala Phe Glu Phe Leu Lys His Ala Trp Lys Ile Ala Gly Lys
            1060             1065             1070

atc aga cta aaa cca act tgt tca ttt tgatttaatc ttaacattat         3362
Ile Arg Leu Lys Pro Thr Cys Ser Phe
        1075             1080

agcaactgca aggtcttgat catctgttag ttgcgtacta acttatgtgt attagtataa  3422

caagaaaaga gaattagaga gatggattct aatccggtgt tgcagtacat cttttctcca  3482

cccgcataaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa                        3522
```

```
<210>        19
<211>        1081
<212>        PRT
```

14

<213>   *Arabidopsis thaliana* ecotype Columbia
<223>   Polypeptide MSH3

<400>   19

Met Gly Lys Gln Lys Gln Gln Thr Ile Ser Arg Phe Phe Ala Pro Lys
 1   5   10   15

Pro Lys Ser Pro Thr His Glu Pro Asn Pro Val Ala Glu Ser Ser Thr
                20   25   30

Pro Pro Pro Lys Ile Ser Ala Thr Val Ser Phe Ser Pro Ser Lys Arg
            35   40   45

Lys Leu Leu Ser Asp His Leu Ala Ala Ala Ser Pro Lys Lys Pro Lys
        50   55   60

Leu Ser Pro His Thr Gln Asn Pro Val Pro Asp Pro Asn Leu His Gln
65   70   75   80

Arg Phe Leu Gln Arg Phe Leu Glu Pro Ser Pro Glu Glu Tyr Val Pro
                85   90   95

Glu Thr Ser Ser Ser Arg Lys Tyr Thr Pro Leu Glu Gln Gln Val Val
            100   105   110

Glu Leu Lys Ser Lys Tyr Pro Asp Val Val Leu Met Val Glu Val Gly
        115   120   125

Tyr Arg Tyr Arg Phe Phe Gly Glu Asp Ala Glu Ile Ala Ala Arg Val
    130   135   140

Leu Gly Ile Tyr Ala His Met Asp His Asn Phe Met Thr Ala Ser Val
145   150   155   160

Pro Thr Phe Arg Leu Asn Phe His Val Arg Arg Leu Val Asn Ala Gly
                165   170   175

Tyr Lys Ile Gly Val Val Lys Gln Thr Glu Thr Ala Ala Ile Lys Ser
            180   185   190

His Gly Ala Asn Arg Thr Gly Pro Phe Phe Arg Gly Leu Ser Ala Leu
        195   200   205

Tyr Thr Lys Ala Thr Leu Glu Ala Ala Glu Asp Ile Ser Gly Gly Cys
    210   215   220

Gly Gly Glu Glu Gly Phe Gly Ser Gln Ser Asn Phe Leu Val Cys Val
225   230   235   240

Val Asp Glu Arg Val Lys Ser Glu Thr Leu Gly Cys Gly Ile Glu Met
                245   250   255

15

Ser Phe Asp Val Arg Val Gly Val Val Gly Val Glu Ile Ser Thr Gly
              260   265   270

Glu Val Val Tyr Glu Glu Phe Asn Asp Asn Phe Met Arg Ser Gly Leu
              275   280   285

Glu Ala Val Ile Leu Ser Leu Ser Pro Ala Glu Leu Leu Leu Gly Gln
         290   295   300

Pro Leu Ser Gln Gln Thr Glu Lys Phe Leu Val Ala Met Ala Gly Pro
305   310   315   320

Thr Ser Asn Val Arg Val Glu Arg Ala Ser Leu Asp Cys Phe Ser Asn
                   325   330   335

Gly Asn Ala Val Asp Glu Val Ile Ser Leu Cys Glu Lys Ile Ser Ala
              340   345   350

Gly Asn Leu Glu Asp Asp Lys Glu Met Lys Leu Glu Ala Ala Glu Lys
         355   360   365

Gly Met Ser Cys Leu Thr Val His Thr Ile Met Asn Met Pro His Leu
    370   375   380

Thr Val Gln Ala Leu Ala Leu Thr Phe Cys His Leu Lys Gln Phe Gly
385   390   395   400

Phe Glu Arg Ile Leu Tyr Gln Gly Ala Ser Phe Arg Ser Leu Ser Ser
                   405   410   415

Asn Thr Glu Met Thr Leu Ser Ala Asn Thr Leu Gln Gln Leu Glu Val
              420   425   430

Val Lys Asn Asn Ser Asp Gly Ser Glu Ser Gly Ser Leu Phe His Asn
         435   440   445

Met Asn His Thr Leu Thr Val Tyr Gly Ser Arg Leu Leu Arg His Trp
    450   455   460

Val Thr His Pro Leu Cys Asp Arg Asn Leu Ile Ser Ala Arg Leu Asp
465   470   475   480

Ala Val Ser Glu Ile Ser Ala Cys Met Gly Ser His Ser Ser Ser Gln
                   485   490   495

Leu Ser Ser Glu Leu Val Glu Glu Gly Ser Glu Arg Ala Ile Val Ser
              500   505   510

Pro Glu Phe Tyr Leu Val Leu Ser Ser Val Leu Thr Ala Met Ser Arg
         515   520   525

Ser Ser Asp Ile Gln Arg Gly Ile Thr Arg Ile Phe His Arg Thr Ala
         530   535   540

Lys Ala Thr Glu Phe Ile Ala Val Met Glu Ala Ile Leu Leu Ala Gly
545   550   555   560

Lys Gln Ile Gln Arg Leu Gly Ile Lys Gln Asp Ser Glu Met Arg Ser
                565   570   575

Met Gln Ser Ala Thr Val Arg Ser Thr Leu Leu Arg Lys Leu Ile Ser
         580   585   590

Val Ile Ser Ser Pro Val Val Asp Asn Ala Gly Lys Leu Leu Ser
         595   600   605

Ala Leu Asn Lys Glu Ala Ala Val Arg Gly Asp Leu Leu Asp Ile Leu
    610   615   620

Ile Thr Ser Ser Asp Gln Phe Pro Glu Leu Ala Glu Ala Arg Gln Ala
625   630   635   640

Val Leu Val Ile Arg Glu Lys Leu Asp Ser Ser Ile Ala Ser Phe Arg
             645   650   655

Lys Lys Leu Ala Ile Arg Asn Leu Glu Phe Leu Gln Val Ser Gly Ile
         660   665   670

Thr His Leu Ile Glu Leu Pro Val Asp Ser Lys Val Pro His Asn Trp
         675   680   685

Val Lys Val Asn Ser Thr Lys Lys Thr Ile Arg Tyr His Pro Pro Glu
    690   695   700

Ile Val Ala Gly Leu Asp Glu Leu Ala Leu Ala Thr Glu His Leu Ala
705   710   715   720

Ile Val Asn Arg Ala Ser Trp Asp Ser Phe Leu Lys Ser Phe Ser Arg
                725   730   735

Tyr Tyr Thr Asp Phe Lys Ala Ala Val Gln Ala Leu Ala Ala Leu Asp
         740   745   750

Cys Leu His Ser Leu Ser Thr Leu Ser Arg Asn Lys Asn Tyr Val Arg
         755   760   765

Pro Glu Phe Val Asp Asp Cys Glu Pro Val Glu Ile Asn Ile Gln Ser
    770   775   780

Gly Arg His Pro Val Leu Glu Thr Ile Leu Gln Asp Asn Phe Val Pro
785   790   795   800

Asn Asp Thr Ile Leu His Ala Glu Gly Glu Tyr Cys Gln Ile Ile Thr
             805   810   815

Gly Pro Asn Met Gly Gly Lys Ser Cys Tyr Ile Arg Gln Val Ala Leu
         820   825   830

Ile Ser Ile Met Ala Gln Val Gly Ser Phe Val Pro Ala Ser Phe Ala
        835   840   845

Lys Leu His Val Leu Asp Gly Val Phe Thr Arg Met Gly Ala Ser Asp
        850   855   860

Ser Ile Gln His Gly Arg Ser Thr Phe Leu Glu Glu Leu Ser Glu Ala
865   870   875   880

Ser His Ile Ile Arg Thr Cys Ser Ser Arg Ser Leu Val Ile Leu Asp
                885   890   895

Glu Leu Gly Arg Gly Thr Ser Thr His Asp Gly Val Ala Ile Ala Tyr
            900   905   910

Ala Thr Leu Gln His Leu Leu Ala Glu Lys Arg Cys Leu Val Leu Phe
        915   920   925

Val Thr His Tyr Pro Glu Ile Ala Glu Ile Ser Asn Gly Phe Pro Gly
        930   935   940

Ser Val Gly Thr Tyr His Val Ser Tyr Leu Thr Leu Gln Lys Asp Lys
945   950   955   960

Gly Ser Tyr Asp His Asp Asp Val Thr Tyr Leu Tyr Lys Leu Val Arg
                965   970   975

Gly Leu Cys Ser Arg Ser Phe Gly Phe Lys Val Ala Gln Leu Ala Gln
            980   985   990

Ile Pro Pro Ser Cys Ile Arg Arg Ala Ile Ser Met Ala Ala Lys Leu
        995   1000   1005

Glu Ala Glu Val Arg Ala Arg Glu Arg Asn Thr Arg Met Gly Glu Pro
    1010   1015   1020

Glu Gly His Glu Glu Pro Arg Gly Ala Glu Glu Ser Ile Ser Ala Leu
1025   1030   1035   1040

Gly Asp Leu Phe Ala Asp Leu Lys Phe Ala Leu Ser Glu Glu Asp Pro
            1045   1050   1055

Trp Lys Ala Phe Glu Phe Leu Lys His Ala Trp Lys Ile Ala Gly Lys
    1060   1065   1070

Ile Arg Leu Lys Pro Thr Cys Ser Phe
        1075   1080

<210>   20
<211>   24
<212>   DNA
<213>   Artificial sequence

```
<220>
<223>      MSH6 specific primer 638 for PCR using cDNA of Arabidopsis
           thaliana ecotype Columbia

<400>      20

tctctaccag gtgacgaaaa accg                                              24


<210>      21
<211>      28
<212>      DNA
<213>      Artificial sequence

<220>
<223>      Primer S81 for PCR using cDNA of Arabidopsis thaliana ecotype
           Columbia

<400>      21

cgtcgccttt agcatcccct tccttcac                                         28


<210>      22
<211>      30
<212>      DNA
<213>      Artificial sequence

<220>
<223>      MSH6 specific primer S823 for PCR using cDNA of Arabidopsis
           thaliana ecotype Columbia

<400>      22

gcttggcgca tctaatagaa tcatgacagg                                       30


<210>      23
<211>      24
<212>      DNA
<213>      Artificial sequence

<220>
<223>      MSH6 specific primer 637 for PCR using cDNA of Arabidopsis
           thaliana ecotype Columbia

<400>      23

gacagcgtca gttcttcaga atgc                                             24


<210>      24
<211>      33
<212>      DNA
```

<213>   Artificial sequence

<220>
<223>   MSH6 specific primer 1S8 for PCR using cDNA of Arabidopsis
               thaliana ecotype Columbia

<400>   24

atcccgggat gcagcgccag agatcgattt tgt   33

<210>   25
<211>   27
<212>   DNA
<213>   Artificial sequence

<220>
<223>   MSH6 specific primer S83 for PCR using cDNA of Arabidopsis
               thaliana ecotype Columbia

<400>   25

cgctatctat ggctgcttcg aattgag   27

<210>   26
<211>   1385
<212>   DNA
<213>   Arabidopsis thaliana ecotype Columbia
<223>   Clone 43

<400>   26

cccgggatgc agcgccagag atcgattttg tctttcttcc aaaaacccac ggcggcgact   60

acgaagggtt tggtttccgg cgatgctgct agcggcgggg gcggcagcgg aggaccacga   120

tttaatgtga aggaagggga tgctaaaggc gacgcttctg tacgttttgc tgtttcgaaa   180

tctgtcgatg aggttagagg aacggatact ccaccggaga aggttccgcg tcgtgtcctg   240

ccgtctggat ttaagccggc tgaatccgcc ggtgatgctt cgtccctgtt ctccaatatt   300

atgcataagt ttgtaaaagt cgatgatcga gattgttctg agagaggag ccgagaagat   360

gttgttccgc tgaatgattc atctctatgt atgaaggcta atgatgttat tcctcaattt   420

cgttccaata atggtaaaac tcaagaaaga aaccatgctt ttagtttcag tgggagagct   480

gaacttagat cagtagaaga tataggagta gatggcgatg ttcctggtcc agaaacacca   540

gggatgcgtc cacgtgcttc tcgcttgaag cgagttctgg aggatgaaat gacttttaag   600

gaggataagg ttcctgtatt ggactctaac aaaaggctga aaatgctcca ggatccggtt   660

```
tgtggagaga agaaagaagt aaacgaagga accaaatttg aatggcttga gtcttctcga      720

atcagggatg ccaatagaag acgtcctgat gatccccttt acgatagaaa gaccttacac      780

ataccacctg atgttttcaa gaaaatgtct gcatcacaaa agcaatattg gagtgttaag      840

agtgaatata tggacattgt gcttttcttt aaagtgggga aattttatga gctgtatgag      900

ctagatgcgg aattaggtca caaggagctt gactggaaga tgaccatgag tggtgtggga      960

aaatgcagac aggttggtat ctctgaaagt gggatagatg aggcagtgca aaagctatta     1020

gctcgtggat ataaagttgg acgaatcgag cagctagaaa catctgacca agcaaaagcc     1080

agaggtgcta atactataat tccaaggaag ctagttcagg tattaactcc atcaacagca     1140

agcgagggaa acatcgggcc tgatgccgtc catcttcttg ctataaaaga gatcaaaatg     1200

gagctacaaa agtgttcaac tgtgtatgga tttgcttttg ttgactgtgc tgccttgagg     1260

ttttgggttg ggtccatcag cgatgatgca tcatgtgctg ctcttggagc gttattgatg     1320

caggtttctc caaaggaagt gttatatgac agtaaagggc tatcaagaga agcacaaaag     1380

gctctaagga aatatacgtt gacagggtct acggcggtac agttggctcc agtaccacaa     1440

gtaatggggg atacagatgc tgctggagtt agaaatataa tagaatctaa cggatacttt     1500

aaaggttctt ctgaatcatg gaactgtgct gttgatggtc taaatgaatg tgatgttgcc     1560

cttagtgctc ttggagagct aattaatcat ctgtctaggc taaagctaga agatgtactt     1620

aagcatgggg atattttttcc ataccaagtt tacaggggtt gtctcagaat tgatggccag     1680

acgatggtaa atcttgagat atttaacaat agctgtgatg gtggtccttc agggaccttg     1740

tacaaatatc ttgataactg tgttagtcca actggtaagc gactcttaag gaattggatc     1800

tgccatccac tcaaagatgt agaaagcatc aataaacggc ttgatgtagt tgaagaattc     1860

acggcaaact cagaaagtat gcaaatcact ggccagtatc tccacaaact tccagactta     1920

gaaagactgc tcggacgcat caagtctagc gttcgatcat cagcctctgt gttgcctgct     1980

cttctgggga aaaaagtgct gaaacaacga gttaaagcat ttgggcaaat tgtgaaaggg     2040

ttcagaagtg gaattgatct gttgttggct ctacagaagg aatcaaatat gatgagtttg     2100

ctttataaac tctgtaaact tcctatatta gtaggaaaaa gcgggctaga gttatttctt     2160

tctcaattcg aagcagccat agatagcg                                        2188
```

<210>   27
<211>   1385
<212>   DNA
<213>   Arabidopsis thaliana ecotype Columbia
<223>   Clone 62

<400>   27

catcagcctc tgtgttgcct gctcttctgg ggaaaaaagt gctgaaacaa cgagttaaag   60

catttgggca aattgtgaaa gggttcagaa gtggaattga tctgttgttg gctctacaga   120

aggaatcaaa tatgatgagt ttgctttata aactctgtaa acttcctata ttagtaggaa   180

aaagcgggct agagttattt ctttctcaat tcgaagcagc catagatagc gactttccaa   240

attatcagaa ccaagatgtg acagatgaaa cgctgaaac tctcacaata cttatcgaac   300

tttttatcga aagagcaact caatggtctg aggtcattca caccataagc tgcctagatg   360

tcctgagatc ttttgcaatc gcagcaagtc tctctgctgg aagcatggcc aggcctgtta   420

ttttttcccga atcagaagct acagatcaga atcagaaac aaaagggcca atacttaaaa   480

tccaaggact atggcatcca tttgcagttg cagccgatgg tcaattgcct gttccgaatg   540

atatactcct tggcgaggct agaagaagca gtggcagcat tcatcctcgg tcattgttac   600

tgacgggacc aaacatgggc ggaaaatcaa ctcttcttcg tgcaacatgt ctggccgtta   660

tctttgccca acttggctgc tacgtgccgt gtgagtcttg cgaaatctcc ctcgtggata   720

ctatcttcac aaggcttggc gcatctgata gaatcatgac aggagagagt acctttttgg   780

tagaatgcac tgagacagcg tcagttcttc agaatgcaac tcaggattca ctagtaatcc   840

ttgacgaact gggcagagga actagtactt tcgatggata cgccattgca tactcggttt   900

ttcgtcacct ggtagagaaa gttcaatgtc ggatgctctt gcaacacat taccaccctc   960

tcaccaagga attcgcgtct cacccacgtg tcacctcgaa acacatggct tgcgcattca   1020

aatcaagatc tgattatcaa ccacgtggtt gtgatcaaga cctagtgttc ttgtaccgtt   1080

taaccgaggg agcttgtcct gagagctacg gacttcaagt ggcactcatg ctggaatac   1140

caaaccaagt ggttgaaaca gcatcaggtg ctgctcaagc catgaagaga tcaattgggg   1200

aaaacttcaa gtcaagtgag ctaagatctg agttctcaag tctgcatgaa gactggctca   1260

agtcattggt gggtatttct cgagtcgccc acaacaatgc ccccattggc gaagatgact   1320

acgacacttt gttttgctta tggcatgaga tcaaatcctc ttactgtgtt cccaaataac   1380

ccggg
<div align="right">1385</div>

```
<210>      28
<211>      34
<212>      DNA
<213>      Artificial sequence

<220>
<223>      MSH6 specific primer 2S8 for PCR using cDNA of Arabidopsis
           thaliana ecotype Columbia

<400>      26
```

atcccgggtt atttgggaac acagtaagag gatt   34

```
<210>      29
<211>      27
<212>      DNA
<213>      Artificial sequence

<220>
<223>      MSH6 specific primer S82 for PCR using cDNA of Arabidopsis
           thaliana ecotype Columbia

<400>      29
```

gcgttcgatc atcagcctct gtgttgc   27

```
<210>      30
<211>      3606
<212>      DNA
<213>      Arabidopsis thaliana ecotype Columbia

<220>
<221>      CDS
<222>      (142)....(3468)
<223>      AtMSH6 full-length cDNA and deduced sequence of the encoded
           polypeptide

<400>      30
```

aaaagttgag ccctgaggag tatcgtttcc gccatttcta cgacgcaagg cgaaaatttt   60

tggcgccaat ctttcccccc tttcgaattc tctcagctca aaacatcgtt tctctctcac   120

tctctctcac aattccaaaa a atg cag cgc cag aga tcg att ttg tct ttc   171
                        Met Gln Arg Gln Arg Ser Ile Leu Ser Phe
                         1   5   10

```
ttc caa aaa ccc acc gcg gcg act acg aag ggt ttg gtt tcc ggc gat       219
Phe Gln Lys Pro Thr Ala Ala Thr Thr Lys Gly Leu Val Ser Gly Asp
                15                  20                  25


gct gct agc ggc ggg ggc ggc agc gga gga cca cga ttt aat gtg aag       267
Ala Ala Ser Gly Gly Gly Gly Ser Gly Gly Pro Arg Phe Asn Val Arg
            30                  35                  40


gaa ggg gat gct aaa ggc gac gct tct gta cgt ttt gct gtt tcg aaa       315
Glu Gly Asp Ala Lys Gly Asp Ala Ser Val Arg Phe Ala Val Ser Lys
            45                  50                  55


tct gtc gat gag gtt aga gga acg gat act cca ccg gag aag gtt ccg       363
Ser Val Asp Glu Val Arg Gly Thr Asp Thr Pro Pro Glu Lys Val Pro
        60                  65                  70


cgt cgt gtc ctg ccg tct gga ttt aag ccg gct gaa tcc gcc gst gat       411
Arg Arg Val Leu Pro Ser Gly Phe Lys Pro Ala Glu Ser Ala Gly Asp
75                  80                  85                  90


gct tcg tcc ctg ttc tcc aat att atg cat aag ttt gta aaa gtc gat       459
Ala Ser Ser Leu Phe Ser Asn Ile Met His Lys Phe Val Lys Val Asp
                95                  100                 105


gat cga gat tgt tct gga gag agg agc cga gaa gat gtt gtt ccg ctg       507
Asp Arg Asp Cys Ser Gly Glu Arg Ser Arg Glu Asp Val Val Pro Leu
                110                 115                 120


aat gat tca tct cta tgt atg aag gct aat gat gtt att cct caa ttt       555
Asn Asp Ser Ser Leu Cys Met Lys Ala Asn Asp Val Ile Pro Gln Phe
            125                 130                 135


cgt tcc aat aat ggt aaa act caa gaa aga aac cat gct ttt agt ttc       603
Arg Ser Asn Asn Gly Lys Thr Gln Glu Arg Asn His Ala Phe Ser Phe
        140                 145                 150


agt ggg aga gct gaa ctt aga tca gta gaa gat ata gga gta gat ggc       651
Ser Gly Arg Ala Glu Leu Arg Ser Val Glu Asp Ile Gly Val Asp Gly
155                 160                 165                 170


gat gtt cct ggt cca gaa aca cca ggg atg cgt cca cgt gct tct cgc       699
Asp Val Pro Gly Pro Glu Thr Pro Gly Met Arg Pro Arg Ala Ser Arg
                175                 180                 185


ttg aag cga gtt ctg gag gat gaa atg act ttt aag gag gat aag gtt       747
Leu Lys Arg Val Leu Glu Asp Glu Met Thr Phe Lys Glu Asp Lys Val
                190                 195                 200


cct gta ttg gac tct aac aaa agg ctg aaa atg ctc cag gat ccg gtt       795
Pro Val Leu Asp Ser Asn Lys Arg Leu Lys Met Leu Gln Asp Pro Val
            205                 210                 215
```

```
tgt gga gag aag aaa gaa gta aac gaa gga acc aaa ttt gaa tgg ctt          843
Cys Gly Glu Lys Lys Glu Val Asn Glu Gly Thr Lys Phe Glu Trp Leu
    220                 225                 230

gag tct tct cga atc agg gat gcc aat aga aga cgt cct gat gat ccc          891
Glu Ser Ser Arg Ile Arg Asp Ala Asn Arg Arg Arg Pro Asp Asp Pro
235                 240                 245                 250

ctt tac gat aga aag acc tta cac ata cca cct gat gtt ttc aag aaa          939
Leu Tyr Asp Arg Lys Thr Leu His Ile Pro Pro Asp Val Phe Lys Lys
                    255                 260                 265

atg tct gca tca caa aag caa tat tgg agt gtt aag agt gaa tat atg          987
Met Ser Ala Ser Gln Lys Gln Tyr Trp Ser Val Lys Ser Glu Tyr Met
                270                 275                 280

gac att gtg ctt ttc ttt aaa gtg ggg aaa ttt tat gag ctg tat gag         1035
Asp Ile Val Leu Phe Phe Lys Val Gly Lys Phe Tyr Glu Leu Tyr Glu
            285                 290                 295

cta gat gcg gaa tta ggt cac aag gag ctt gac tgg aag atg acc atg         1083
Leu Asp Ala Glu Leu Gly His Lys Glu Leu Asp Trp Lys Met Thr Met
        300                 305                 310

agt ggt gtg gga aaa tgc aga cag gtt ggt atc tct gaa agt ggg ata         1131
Ser Gly Val Gly Lys Cys Arg Gln Val Gly Ile Ser Glu Ser Gly Ile
315                 320                 325                 330

gat gag gca gtg caa aag cta tta gct cgt gga tat aaa gtt gga cga         1179
Asp Glu Ala Val Gln Lys Leu Leu Ala Arg Gly Tyr Lys Val Gly Arg
                    335                 340                 345

atc gag cag cta gaa aca tct gac caa gca aaa gcc aga ggt gct aat         1227
Ile Glu Gln Leu Glu Thr Ser Asp Gln Ala Lys Ala Arg Gly Ala Asn
                350                 355                 360

act ata att cca agg aag cta gtt cag gta tta act cca tca aca gca         1275
Thr Ile Ile Pro Arg Lys Leu Val Gln Val Leu Thr Pro Ser Thr Ala
            365                 370                 375

agc gag gga aac atc ggg cct gat gcc gtc cat ctt ctt gct ata aaa         1323
Ser Glu Gly Asn Ile Gly Pro Asp Ala Val His Leu Leu Ala Ile Lys
        380                 385                 390

gag atc aaa atg gag cta caa aag tgt tca act gtg tat gga ttt gct         1371
Glu Ile Lys Met Glu Leu Gln Lys Cys Ser Thr Val Tyr Gly Phe Ala
395                 400                 405                 410

ttt gtt gac tgt gct gcc ttg agg ttt tgg gtt ggg tcc atc agc gat         1419
Phe Val Asp Cys Ala Ala Leu Arg Phe Trp Val Gly Ser Ile Ser Asp
            415                 420                 425
```

25

```
gat gca tca tgt gct gct ctt gga gcg tta ttg atg cag gtt tct cca        1467
Asp Ala Ser Cys Ala Ala Leu Gly Ala Leu Leu Met Gln Val Ser Pro
            430                 435                 440

aag gaa gtg tta tat gac agt aaa ggg cta tca aga gaa gca caa aag        1515
Lys Glu Val Leu Tyr Asp Ser Lys Gly Leu Ser Arg Glu Ala Gln Lys
            445                 450                 455

gct cta agg aaa tat acg ttg aca ggg tct acg gcg gta cag ttg gct       1563
Ala Leu Arg Lys Tyr Thr Leu Thr Gly Ser Thr Ala Val Gln Leu Ala
            460                 465                 470

cca gta cca caa gta atg ggg gat aca gat gct gct gga gtt aga aat       1611
Pro Val Pro Gln Val Met Gly Asp Thr Asp Ala Ala Gly Val Arg Asn
475                 480                 485                 490

ata ata gaa tct aac gga tac ttt aaa ggt tct tct gaa tca tgg aac       1659
Ile Ile Glu Ser Asn Gly Tyr Phe Lys Gly Ser Ser Glu Ser Trp Asn
                495                 500                 505

tgt gct gtt gat ggt cta aat gaa tgt gat gtt gcc ctt agt gct ctt       1707
Cys Ala Val Asp Gly Leu Asn Glu Cys Asp Val Ala Leu Ser Ala Leu
            510                 515                 520

gga gag cta att aat cat ctg tct agg cta aag cta gaa gat gta ctt       1755
Gly Glu Leu Ile Asn His Leu Ser Arg Leu Lys Leu Glu Asp Val Leu
            525                 530                 535

aag cat ggg gat att ttt cca tac caa gtt tac agg ggt tgt ctc aga       1803
Lys His Gly Asp Ile Phe Pro Tyr Gln Val Tyr Arg Gly Cys Leu Arg
            540                 545                 550

att gat ggc cag acg atg gta aat ctt gag ata ttt aac aat agc tgt       1851
Ile Asp Gly Gln Thr Met Val Asn Leu Glu Ile Phe Asn Asn Ser Cys
555                 560                 565                 570

gat ggt ggt cct tca ggg acc ttg tac aaa tat ctt gat aac tgt gtt       1899
Asp Gly Gly Pro Ser Gly Thr Leu Tyr Lys Tyr Leu Asp Asn Cys Val
                575                 580                 585

agt cca act ggt aag cga ctc tta agg aat tgg atc tgc cat cca ctc       1947
Ser Pro Thr Gly Lys Arg Leu Leu Arg Asn Trp Ile Cys His Pro Leu
            590                 595                 600

aaa gat gta gaa agc atc aat aaa cgg ctt gat gta gtt gaa gaa ttc       1995
Lys Asp Val Glu Ser Ile Asn Lys Arg Leu Asp Val Val Glu Glu Phe
            605                 610                 615

acg gca aac tca gaa agt atg caa atc act ggc cag tat ctc cac aaa       2043
Thr Ala Asn Ser Glu Ser Met Gln Ile Thr Gly Gln Tyr Leu His Lys
            620                 625                 630
```

```
ctt cca gac tta gaa aga ctg ctc gga cgc atc aag tct agc gtt cga      2091
Leu Pro Asp Leu Glu Arg Leu Leu Gly Arg Ile Lys Ser Ser Val Arg
635             640             645             650

tca tca gcc tct gtg ttg cct gct ctt ctg ggg aaa aaa gtg ctg aaa      2139
Ser Ser Ala Ser Val Leu Pro Ala Leu Leu Gly Lys Lys Val Leu Lys
        655             660             665

caa cga gtt aaa gca ttt ggg caa att gtg aaa ggg ttc aga agt gga      2187
Gln Arg Val Lys Ala Phe Gly Gln Ile Val Lys Gly Phe Arg Ser Gly
        670             675             680

att gat ctg ttg ttg gct cta cag aag gaa tca aat atg atg agt ttg      2235
Ile Asp Leu Leu Leu Ala Leu Gln Lys Glu Ser Asn Met Met Ser Leu
        685             690             695

ctt tat aaa ctc tgt aaa ctt cct ata tta gta gga aaa agc ggg cta      2283
Leu Tyr Lys Leu Cys Lys Leu Pro Ile Leu Val Gly Lys Ser Gly Leu
        700             705             710

gag tta ttt ctt tct caa ttc gaa gca gcc ata gat agc gac ttt cca      2331
Glu Leu Phe Leu Ser Gln Phe Glu Ala Ala Ile Asp Ser Asp Phe Pro
715             720             725             730

aat tat cag aac caa gat gtg aca gat gaa aac gct gaa act ctc aca      2379
Asn Tyr Gln Asn Gln Asp Val Thr Asp Glu Asn Ala Glu Thr Leu Thr
            735             740             745

ata ctt atc gaa ctt ttt atc gaa aga gca act caa tgg tct gag gtc      2427
Ile Leu Ile Glu Leu Phe Ile Glu Arg Ala Thr Gln Trp Ser Glu Val
            750             755             760

att cac acc ata agc tgc cta gat gtc ctg aga tct ttt gca atc gca      2475
Ile His Thr Ile Ser Cys Leu Asp Val Leu Arg Ser Phe Ala Ile Ala
            765             770             775

gca agt ctc tct gct gga agc atg gcc agg cct gtt att ttt ccc gaa      2523
Ala Ser Leu Ser Ala Gly Ser Met Ala Arg Pro Val Ile Phe Pro Glu
        780             785             790

tca gaa gct aca gat cag aat cag aaa aca aaa ggg cca ata ctt aaa      2571
Ser Glu Ala Thr Asp Gln Asn Gln Lys Thr Lys Gly Pro Ile Leu Lys
795             800             805             810

atc caa gga cta tgg cat cca ttt gca gtt gca gcc gat ggt caa ttg      2619
Ile Gln Gly Leu Trp His Pro Phe Ala Val Ala Ala Asp Gly Gln Leu
            815             820             825

cct gtt ccg aat gat ata ctc ctt ggc gag gct aga aga agc agt ggc      2667
Pro Val Pro Asn Asp Ile Leu Leu Gly Glu Ala Arg Arg Ser Ser Gly
        830             835             840
```

27

```
agc att cat cct cgg tca ttg tta ctg acg gga cca aac atg ggc gga    2715
Ser Ile His Pro Arg Ser Leu Leu Leu Thr Gly Pro Asn Met Gly Gly
        845                 850                 855

aaa tca act ctt ctt cgt gca aca tgt ctg gcc gtt atc ttt gcc caa    2763
Lys Ser Thr Leu Leu Arg Ala Thr Cys Leu Ala Val Ile Phe Ala Gln
        860                 865                 870

ctt ggc tgc tac gtg ccg tgt gag tct tgc gaa atc tcc ctc gtg gat    2811
Leu Gly Cys Tyr Val Pro Cys Glu Ser Cys Glu Ile Ser Leu Val Asp
875                 880                 885                 890

act atc ttc aca agg ctt ggc gca tct gat aga atc atg aca gga gag    2859
Thr Ile Phe Thr Arg Leu Gly Ala Ser Asp Arg Ile Met Thr Gly Glu
                895                 900                 905

agt acc ttt ttg gta gaa tgc act gag aca gcg tca gtt ctt cag aat    2907
Ser Thr Phe Leu Val Glu Cys Thr Glu Thr Ala Ser Val Leu Gln Asn
        910                 915                 920

gca act cag gat tca cta gta atc ctt gac gaa ctg ggc aga gga act    2955
Ala Thr Gln Asp Ser Leu Val Ile Leu Asp Glu Leu Gly Arg Gly Thr
        925                 930                 935

agt act ttc gat gga tac gcc att gca tac tcg gtt ttt cgt cac ctg    3003
Ser Thr Phe Asp Gly Tyr Ala Ile Ala Tyr Ser Val Phe Arg His Leu
        940                 945                 950

gta gag aaa gtt caa tgt cgg atg ctc ttt gca aca cat tac cac cct    3051
Val Glu Lys Val Gln Cys Arg Met Leu Phe Ala Thr His Tyr His Pro
955                 960                 965                 970

ctc acc aag gaa ttc gcg tct cac cca cgt gtc acc tcg aaa cac atg    3099
Leu Thr Lys Glu Phe Ala Ser His Pro Arg Val Thr Ser Lys His Met
                975                 980                 985

gct tgc gca ttc aaa tca aga tct gat tat caa cca cgt ggt tgt gat    3147
Ala Cys Ala Phe Lys Ser Arg Ser Asp Tyr Gln Pro Arg Gly Cys Asp
                990                 995                 1000

caa gac cta gtg ttc ttg tac cgt tta acc gag gga gct tgt cct gag    3195
Gln Asp Leu Val Phe Leu Tyr Arg Leu Thr Glu Gly Ala Cys Pro Glu
        1005                1010                1015

agc tac gga ctt caa gtg gca ctc atg gct gga ata cca aac caa gtg    3243
Ser Tyr Gly Leu Gln Val Ala Leu Met Ala Gly Ile Pro Asn Gln Val
        1020                1025                1030

gtt gaa aca gca tca ggt gct gct caa gcc atg aag aga tca att ggg    3291
Val Glu Thr Ala Ser Gly Ala Ala Gln Ala Met Lys Arg Ser Ile Gly
1035                1040                1045                1050
```

```
gga aac ttc aag tca agt gag cta aga tct gag ttc tca agt ctg cat       3339
Glu Asn Phe Lys Ser Ser Glu Leu Arg Ser Glu Phe Ser Ser Leu His
        1055             1060             1065

gaa gac tgg ctc aag tca ttg gtg ggt att tct cga gtc gcc cac aac       3387
Glu Asp Trp Leu Lys Ser Leu Val Gly Ile Ser Arg Val Ala His Asn
        1070             1075             1080

aat gcc ccc att ggc gaa gat gac tac gac act ttg ttt tgc tta tgg       3435
Asn Ala Pro Ile Gly Glu Asp Asp Tyr Asp Thr Leu Phe Cys Leu Trp
        1085             1090             1095

cat gag atc aaa tcc tct tac tgt gtt ccc aaa taaatggcta                3478
His Glu Ile Lys Ser Ser Tyr Cys Val Pro Lys
        1100             1105

tgacataaca ctatctgaag ctcgttaagt cttttgcctc tctgatgttt attcctctta     3538

aaaaatgctt atatatcaaa aaattgtttc ctcgattaaa aaaaaaaaaa aaaaaaaaaa     3598

aaaaaaaa                                                              3606


<210>      31
<211>      1109
<212>      PRT
<213>      Arabidopsis thaliana ecotype Columbia
<223>      Polypeptide MSH6

<400>      31

Met Gln Arg Gln Arg Ser Ile Leu Ser Phe Phe Gln Lys Pro Thr Ala
1               5                   10                  15

Ala Thr Thr Lys Gly Leu Val Ser Gly Asp Ala Ala Ser Gly Gly Gly
            20                  25                  30

Gly Ser Gly Gly Pro Arg Phe Asn Val Arg Glu Gly Asp Ala Lys Gly
            35                  40                  45

Asp Ala Ser Val Arg Phe Ala Val Ser Lys Ser Val Asp Glu Val Arg
        50                  55                  60

Gly Thr Asp Thr Pro Pro Glu Lys Val Pro Arg Arg Val Leu Pro Ser
65                  70                  75                  80

Gly Phe Lys Pro Ala Glu Ser Ala Gly Asp Ala Ser Ser Leu Phe Ser
                85                  90                  95

Asn Ile Met His Lys Phe Val Lys Val Asp Asp Arg Asp Cys Ser Gly
                100                 105                 110

Glu Arg Ser Arg Glu Asp Val Val Pro Leu Asn Asp Ser Ser Leu Cys
        115                 120                 125
```

```
Met Lys Ala Asn Asp Val Ile Pro Gln Phe Arg Ser Asn Asn Gly Lys
    130             135             140

Thr Gln Glu Arg Asn His Ala Phe Ser Phe Ser Gly Arg Ala Glu Leu
145             150             155             160

Arg Ser Val Glu Asp Ile Gly Val Asp Gly Asp Val Pro Gly Pro Glu
                165             170             175

Thr Pro Gly Met Arg Pro Arg Ala Ser Arg Leu Lys Arg Val Leu Glu
            180             185             190

Asp Glu Met Thr Phe Lys Glu Asp Lys Val Pro Val Leu Asp Ser Asn
        195             200             205

Lys Arg Leu Lys Met Leu Gln Asp Pro Val Cys Gly Glu Lys Lys Glu
    210             215             220

Val Asn Glu Gly Thr Lys Phe Glu Trp Leu Glu Ser Ser Arg Ile Arg
225             230             235             240

Asp Ala Asn Arg Arg Arg Pro Asp Asp Pro Leu Tyr Asp Arg Lys Thr
            245             250             255

Leu His Ile Pro Pro Asp Val Phe Lys Lys Met Ser Ala Ser Gln Lys
            260             265             270

Gln Tyr Trp Ser Val Lys Ser Glu Tyr Met Asp Ile Val Leu Phe Phe
        275             280             285

Lys Val Gly Lys Phe Tyr Glu Leu Tyr Glu Leu Asp Ala Glu Leu Gly
    290             295             300

His Lys Glu Leu Asp Trp Lys Met Thr Met Ser Gly Val Gly Lys Cys
305             310             315             320

Arg Gln Val Gly Ile Ser Glu Ser Gly Ile Asp Glu Ala Val Gln Lys
            325             330             335

Leu Leu Ala Arg Gly Tyr Lys Val Gly Arg Ile Glu Gln Leu Glu Thr
            340             345             350

Ser Asp Gln Ala Lys Ala Arg Gly Ala Asn Thr Ile Ile Pro Arg Lys
        355             360             365

Leu Val Gln Val Leu Thr Pro Ser Thr Ala Ser Glu Gly Asn Ile Gly
    370             375             380

Pro Asp Ala Val His Leu Leu Ala Ile Lys Glu Ile Lys Met Glu Leu
385             390             395             400

Gln Lys Cys Ser Thr Val Tyr Gly Phe Ala Phe Val Asp Cys Ala Ala
            405             410             415
```

Leu Arg Phe Trp Val Gly Ser Ile Ser Asp Asp Ala Ser Cys Ala Ala
        420   425   430

Leu Gly Ala Leu Leu Met Gln Val Ser Pro Lys Glu Val Leu Tyr Asp
        435   440   445

Ser Lys Gly Leu Ser Arg Glu Ala Gln Lys Ala Leu Arg Lys Tyr Thr
    450   455   460

Leu Thr Gly Ser Thr Ala Val Gln Leu Ala Pro Val Pro Gln Val Met
465   470   475   480

Gly Asp Thr Asp Ala Ala Gly Val Arg Asn Ile Ile Glu Ser Asn Gly
                485   490   495

Tyr Phe Lys Gly Ser Ser Glu Ser Trp Asn Cys Ala Val Asp Gly Leu
            500   505   510

Asn Glu Cys Asp Val Ala Leu Ser Ala Leu Gly Glu Leu Ile Asn His
        515   520   525

Leu Ser Arg Leu Lys Leu Glu Asp Val Leu Lys His Gly Asp Ile Phe
    530   535   540

Pro Tyr Gln Val Tyr Arg Gly Cys Leu Arg Ile Asp Gly Gln Thr Met
545   550   555   560

Val Asn Leu Glu Ile Phe Asn Asn Ser Cys Asp Gly Gly Pro Ser Gly
                565   570   575

Thr Leu Tyr Lys Tyr Leu Asp Asn Cys Val Ser Pro Thr Gly Lys Arg
            580   585   590

Leu Leu Arg Asn Trp Ile Cys His Pro Leu Lys Asp Val Glu Ser Ile
    595   600   605

Asn Lys Arg Leu Asp Val Val Glu Glu Phe Thr Ala Asn Ser Glu Ser
    610   615   620

Met Gln Ile Thr Gly Gln Tyr Leu His Lys Leu Pro Asp Leu Glu Arg
625   630   635   640

Leu Leu Gly Arg Ile Lys Ser Ser Val Arg Ser Ser Ala Ser Val Leu
            645   650   655

Pro Ala Leu Leu Gly Lys Lys Val Leu Lys Gln Arg Val Lys Ala Phe
        660   665   670

Gly Gln Ile Val Lys Gly Phe Arg Ser Gly Ile Asp Leu Leu Leu Ala
        675   680   685

Leu Gln Lys Glu Ser Asn Met Met Ser Leu Leu Tyr Lys Leu Cys Lys
        690   695   700

Leu Pro Ile Leu Val Gly Lys Ser Gly Leu Glu Leu Phe Leu Ser Gln
705   710   715   720

Phe Glu Ala Ala Ile Asp Ser Asp Phe Pro Asn Tyr Gln Asn Gln Asp
                    725   730   735

Val Thr Asp Glu Asn Ala Glu Thr Leu Thr Ile Leu Ile Glu Leu Phe
             740   745   750

Ile Glu Arg Ala Thr Gln Trp Ser Glu Val Ile His Thr Ile Ser Cys
         755   760   765

Leu Asp Val Leu Arg Ser Phe Ala Ile Ala Ala Ser Leu Ser Ala Gly
    770   775   780

Ser Met Ala Arg Pro Val Ile Phe Pro Glu Ser Glu Ala Thr Asp Gln
785   790   795   800

Asn Gln Lys Thr Lys Gly Pro Ile Leu Lys Ile Gln Gly Leu Trp His
             805   810   815

Pro Phe Ala Val Ala Ala Asp Gly Gln Leu Pro Val Pro Asn Asp Ile
             820   825   830

Leu Leu Gly Glu Ala Arg Arg Ser Ser Gly Ser Ile His Pro Arg Ser
         835   840   845

Leu Leu Leu Thr Gly Pro Asn Met Gly Gly Lys Ser Thr Leu Leu Arg
    850   855   860

Ala Thr Cys Leu Ala Val Ile Phe Ala Gln Leu Gly Cys Tyr Val Pro
865   870   875   880

Cys Glu Ser Cys Glu Ile Ser Leu Val Asp Thr Ile Phe Thr Arg Leu
             885   890   895

Gly Ala Ser Asp Arg Ile Met Thr Gly Glu Ser Thr Phe Leu Val Glu
             900   905   910

Cys Thr Glu Thr Ala Ser Val Leu Gln Asn Ala Thr Gln Asp Ser Leu
         915   920   925

Val Ile Leu Asp Glu Leu Gly Arg Gly Thr Ser Thr Phe Asp Gly Tyr
    930   935   940

Ala Ile Ala Tyr Ser Val Phe Arg His Leu Val Glu Lys Val Gln Cys
945   950   955   960

Arg Met Leu Phe Ala Thr His Tyr His Pro Leu Thr Lys Glu Phe Ala
             965   970   975

Ser His Pro Arg Val Thr Ser Lys His Met Ala Cys Ala Phe Lys Ser
         980   985   990

Arg Ser Asp Tyr Gln Pro Arg Gly Cys Asp Gln Asp Leu Val Phe Leu
        995   1000   1005

Tyr Arg Leu Thr Glu Gly Ala Cys Pro Glu Ser Tyr Gly Leu Gln Val
    1010   1015   1020

Ala Leu Met Ala Gly Ile Pro Asn Gln Val Val Glu Thr Ala Ser Gly
1025   1030   1035   1040

Ala Ala Gln Ala Met Lys Arg Ser Ile Gly Glu Asn Phe Lys Ser Ser
            1045   1050   1055

Glu Leu Arg Ser Glu Phe Ser Ser Leu His Glu Asp Trp Leu Lys Ser
            1060   1065   1070

Leu Val Gly Ile Ser Arg Val Ala His Asn Asn Ala Pro Ile Gly Glu
        1075   1080   1085

Asp Asp Tyr Asp Thr Leu Phe Cys Leu Trp His Glu Ile Lys Ser Ser
    1090   1095   1100

Tyr Cys Val Pro Lys
1105

<210>   32
<211>   24
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of ATHGENEA
            microsatellite

<400>   32

accatgcata gcttaaactt cttg   24

<210>   33
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of ATHGENEA
            microsatellite

<400>   33

acataaccac aaatagggt gc   22

```
<210>        34
<211>        18
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Forward primer DMCIN-A for PCR on genomic DNA of Arabidopsis
             thaliana ssp. Landsberg erecta "Ler"

<400>        34

gaagcgatat tgttcgtg                                                      18


<210>        35
<211>        18
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Reverse primer DMCIN-B for PCR on genomic DNA of Arabidopsis
             thaliana ssp. Landsberg erecta "Ler"

<400>        35

agattgcgag aacattcc                                                      18


<210>        36
<211>        31
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Forward primer DMCIN-1 for PCR on genomic DNA of Arabidopsis
             thaliana ssp. Landsberg erecta "Ler"

<400>        36

acgcgtcgac tcagctatga gattactcgt g                                      31


<210>        37
<211>        29
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Reverse primer DMCIN-2 for PCR on genomic DNA of Arabidopsis
             thaliana ssp. Landsberg erecta "Ler"

<400>        37

gctctagatt tctcgctcta agactctct                                         29
```

```
<210>        38
<211>        32
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Forward primer DMCIN-3 for PCR on genomic DNA of Arabidopsis
             thaliana ssp. Landsberg erecta "Ler"

<400>        38

gctctagagc ttctcttaag taagtgattg at                                    32


<210>        39
<211>        48
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Reverse primer DMCIN-4 for PCR on genomic DNA of Arabidopsis
             thaliana ssp. Landsberg erecta "Ler"

<400>        39

tcccccgggc tcgagagatc tccatggttt cttcagctct atgaatcc                   48


<210>        40
<211>        26
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Forward primer DMC1a for PCR on genomic DNA of Arabidopsis
             thaliana ssp. Landsberg erecta "Ler"

<400>        40

acgcgtcgac gaattcgcaa gtgggg                                           26


<210>        41
<211>        38
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Reverse primer DMC1b for PCR on genomic DNA of Arabidopsis
             thaliana ssp. Landsberg erecta "Ler"

<400>        41
```

tccatggaga tctcccgggt accgatttgc ttcgaggg   38


```
<210>        42
<211>        20
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Forward primer for PCR amplification of ATEAT1 SSLP marker in
             Arabidopsis thaliana subspecies

<400>        42
```

gccactgcgt gaatgatatg   20


```
<210>        43
<211>        22
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Reverse primer for PCR amplification of ATEAT1 SSLP marker in
             Arabidopsis thaliana subspecies

<400>        43
```

cgaacagcca acattaattc cc   22


```
<210>        44
<211>        18
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Forward primer for PCR amplification of NGA63 SSLP marker in
             Arabidopsis thaliana subspecies

<400>        44
```

aaccaaggca cagaagcg   18


```
<210>        45
<211>        18
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Reverse primer for PCR amplification of NGA63 SSLP marker in
             Arabidopsis thaliana subspecies
```

<400>   45

acccaagtga tcgccacc   18

<210>   46
<211>   21
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA248 SSLP marker in
             Arabidopsis thaliana subspecies

<400>   46

taccgaacca aaacacaaag g   21

<210>   47
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA248 SSLP marker in
             Arabidopsis thaliana subspecies

<400>   47

tctgtatctc ggtgaattct cc   22

<210>   48
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA128 SSLP marker in
             Arabidopsis thaliana subspecies

<400>   48

ggtctgttga tgtcgtaagt cg   22

<210>   49
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>

<223>   Reverse primer for PCR amplification of NGA128 SSLP marker in
            Arabidopsis thaliana subspecies

<400>   49

atcttgaaac ctttagggag gg   22

<210>   50
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA280 SSLP marker in
            Arabidopsis thaliana subspecies

<400>   50

ctgatctcac ggacaatagt gc   22

<210>   51
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA280 SSLP marker in
            Arabidopsis thaliana subspecies

<400>   51

ggctccataa aaagtgcacc   20

<210>   52
<211>   21
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA111 SSLP marker in
            Arabidopsis thaliana subspecies

<400>   52

ctccagttgg aagctaaagg g   21

<210>   53
<211>   21
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA111 SSLP marker in
               Arabidopsis thaliana subspecies

<400>   53

tgttttttag gacaaatggc g   21

<210>   54
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA168 SSLP marker in
               Arabidopsis thaliana subspecies

<400>   54

ccttcacatc caaaacccac   20

<210>   55
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA168 SSLP marker in
               Arabidopsis thaliana subspecies

<400>   55

gcacataccc acaaccagaa   20

<210>   56
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA1126 SSLP marker
               in Arabidopsis thaliana subspecies

<400>   56

cgctacgctt ttcggtaaag   20

```
<210>        57
<211>        20
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Reverse primer for PCR amplification of NGA1126 SSLP marker
             in Arabidopsis thaliana subspecies

<400>        57


gcacagtcca agtcacaacc                                                     20



<210>        58
<211>        20
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Forward primer for PCR amplification of NGA361 SSLP marker in
             Arabidopsis thaliana subspecies

<400>        58


aaagagatga gaatttggac                                                     20



<210>        59
<211>        23
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Reverse primer for PCR amplification of NGA361 SSLP marker in
             Arabidopsis thaliana subspecies

<400>        59


acatatcaat atattaaagt agc                                                 23



<210>        60
<211>        18
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Forward primer for PCR amplification of NGA168 SSLP marker in
             Arabidopsis thaliana subspecies

<400>        60
```

tcgtctactg cactgccg   18


```
<210>          61
<211>          22
<212>          DNA
<213>          Artificial sequence

<220>
<223>          Reverse primer for PCR amplification of NGA168 SSLP marker in
               Arabidopsis thaliana subspecies

<400>          61
```

gaggacatgt ataggagcct cg   22


```
<210>          62
<211>          20
<212>          DNA
<213>          Artificial sequence

<220>
<223>          Forward primer for PCR amplification of AthBIO2 SSLP marker
               in Arabidopsis thaliana subspecies

<400>          62
```

tgacctcctc ttccatggag   20


```
<210>          63
<211>          22
<212>          DNA
<213>          Artificial sequence

<220>
<223>          Reverse primer for PCR amplification of AthBIO2 SSLP marker
               in Arabidopsis thaliana subspecies

<400>          63
```

ttaacagaaa cccaaagctt tc   22


```
<210>          64
<211>          21
<212>          DNA
<213>          Artificial sequence

<220>
<223>          Forward primer for PCR amplification of AthUBIQUE SSLP marker
               in Arabidopsis thaliana subspecies
```

<400>   64

aggcaaatgt ccatttcatt g   21

<210>   65
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of AthUBIQUE SSLP marker
             in Arabidopsis thaliana subspecies

<400>   65

acgacatggc agatttctcc   20

<210>   66
<211>   21
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA172 SSLP marker in
             Arabidopsis thaliana subspecies

<400>   66

agctgcttcc ttatagcgtc c   21

<210>   67
<211>   19
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA172 SSLP marker in
             Arabidopsis thaliana subspecies

<400>   67

catccgaatg ccattgttc   19

<210>   68
<211>   21
<212>   DNA
<213>   Artificial sequence

<220>

<223>   Forward primer for PCR amplification of NGA126 SSLP marker in
            Arabidopsis thaliana subspecies

<400>   63

gaaaaaacgc tactttcgtg g   21

<210>   69
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA126 SSLP marker in
            Arabidopsis thaliana subspecies

<400>   69   .

caagagcaat atcaagagca gc   22

<210>   70
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA162 SSLP marker in
            Arabidopsis thaliana subspecies

<400>   70

catgcaattt gcatctgagg   20

<210>   71
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA162 SSLP marker in
            Arabidopsis thaliana subspecies

<400>   71

ctctgtcact cttttcctct gg   22

<210>   72
<211>   21
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA6 SSLP marker in
               Arabidopsis thaliana subspecies

<400>   72

tggatttctt cctctcttca c   21

<210>   73
<211>   21
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA6 SSLP marker in
               Arabidopsis thaliana subspecies

<400>   73

atggagaagc ttacactgat c   21

<210>   74
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA12 SSLP marker in
               Arabidopsis thaliana subspecies

<400>   74

aatgttgtcc tcccctcctc   20

<210>   75
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA12 SSLP marker in
               Arabidopsis thaliana subspecies

<400>   75

tgatgctctc tgaaacaaga gc   22

44

```
<210>        76
<211>        21
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Forward primer for PCR amplification of NGA8 SSLP marker in
             Arabidopsis thaliana subspecies

<400>        76


gagggcaaat ctttatttcg g                                            21



<210>        77
<211>        22
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Reverse primer for PCR amplification of NGA8 SSLP marker in
             Arabidopsis thaliana subspecies

<400>        77


tggctttcgt ttataaacat cc                                           22



<210>        78
<211>        21
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Forward primer for PCR amplification of NGA1107 SSLP marker
             in Arabidopsis thaliana subspecies

<400>        78


gcgaaaaaac aaaaaaatcc a                                            21


<210>        79
<211>        21
<212>        DNA
<213>        Artificial sequence

<220>
<223>        Reverse primer for PCR amplification of NGA1107 SSLP marker
             in Arabidopsis thaliana subspecies

<400>        79
```

45

cgacgaatcg acagaattag g   21

<210>	80
<211>	21
<212>	DNA
<213>	Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA225 SSLP marker in
          Arabidopsis thaliana subspecies

<400>   80

gaaatccaaa tcccagagag g   21

<210>	81
<211>	22
<212>	DNA
<213>	Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA225 SSLP marker in
          Arabidopsis thaliana subspecies

<400>   81

tctccccact agttttgtgt cc   22

<210>	82
<211>	19
<212>	DNA
<213>	Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA249 SSLP marker in
          Arabidopsis thaliana subspecies

<400>   82

taccgtcaat ttcatcgcc   19

<210>	83
<211>	22
<212>	DNA
<213>	Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA249 SSLP marker in
          Arabidopsis thaliana subspecies

<400>   83

ggatccctaa ctgtaaaatc cc   22

<210>   84
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of CA72 SSLP marker in
               Arabidopsis thaliana subspecies

<400>   84

aatcccagta accaaacaca ca   22

<210>   85
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of CA72 SSLP marker in
               Arabidopsis thaliana subspecies

<400>   85

cccagtctaa ccacgaccac   20

<210>   86
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA151 SSLP marker in
               Arabidopsis thaliana subspecies

<400>   86

gttttgggaa gttttgctgg   20

<210>   87
<211>   24
<212>   DNA
<213>   Artificial sequence

<220>

<223>   Reverse primer for PCR amplification of NGA151 SSLP marker in
                Arabidopsis thaliana subspecies

<400>   87

cagtctaaaa gcgagagtat gatg   24

<210>   88
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA106 SSLP marker in
                Arabidopsis thaliana subspecies

<400>   88

gttatggagt ttctagggca cg   22

<210>   89
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA106 SSLP marker in
                Arabidopsis thaliana subspecies

<400>   89

tgccccattt tgttcttctc   20

<210>   90
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA139 SSLP marker in
                Arabidopsis thaliana subspecies

<400>   90

agagctacca gatccgatgg   20

<210>   91
<211>   21
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA139 SSLP marker in
             Arabidopsis thaliana subspecies

<400>   91

ggtttcgttt cactatccag g   21

<210>   92
<211>   22
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of NGA76 SSLP marker in
             Arabidopsis thaliana subspecies

<400>   92

ggagaaaatg tcactctcca cc   22

<210>   93
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Reverse primer for PCR amplification of NGA76 SSLP marker in
             Arabidopsis thaliana subspecies

<400>   93

aggcatggga gacatttacg   20

<210>   94
<211>   20
<212>   DNA
<213>   Artificial sequence

<220>
<223>   Forward primer for PCR amplification of ATHSO191 SSLP marker
             in Arabidopsis thaliana subspecies

<400>   94

ctccaccaat catgcaaatg   20

```
<210>         95
<211>         21
<212>         DNA
<213>         Artificial sequence

<220>
<223>         Reverse primer for PCR amplification of ATHSO191 SSLP marker
              in Arabidopsis thaliana subspecies

<400>         95

tgatgttgat ggagatggtc a                                              21


<210>         96
<211>         22
<212>         DNA
<213>         Artificial sequence

<220>
<223>         Forward primer for PCR amplification of NGA129 SSLP marker in
              Arabidopsis thaliana subspecies

<400>         96

tcaggaggaa ctaaagtgag gg                                             22


<210>         97
<211>         22
<212>         DNA
<213>         Artificial sequence

<220>
<223>         Reverse primer for PCR amplification of NGA129 SSLP marker in
              Arabidopsis thaliana subspecies

<400>         97

cacactgaag atggtcttga gg                                             22


<210>         98
<211>         8062
<212>         DNA
<213>         Arabidopsis thaliana ecotype Columbia

<220>
<223>         Genomic DNA sequence of AtMSH6

<400>         97

tttttggtt gctaacaata aaggtatacg gttttatgtc atcaatataa ctatatataa    60
```

```
aagaaatgaa agatatatat tgttttttca tttatcaaac aaaacaacaa gacttttttt      120

ttactttta cattggtcaa caaaatacaa gataaacgac atcgtttaat catttcccaa      180

ttttacccct aagtttaaca cctagaacct tctccatctt cgcaagcaca gcctgattag      240

gaacagcttt accattctca tattcctgaa ctacctgagt cctctcattg atctgtttcg      300

ccaaatccgc ttgtgacatc ttcttctcca atctcgcttt ctgtatcatc aacctcacct      360

ctgctttcac acgatccatc gccgcaggct ctgtttcttc ttccagcttc ttcgtgttaa      420

tcaccggaac cgccgtagat ttcccctttt tgttcgaacc ggcatcgaat ttcttaaccg      480

tttgaaccgc gacaccgttt ctcagagctg cgttaaccgc ttcggatcg cgtaggtctt      540

ggctcttttg ttttgatttg tggagaacta ctggttccca gtcttgtgtt actgctcctg      600

ggtatctgct cggcatcgtc gatgaattga gagaaaggaa caacgcgaaa attttattaa      660

tctgagtttt gaaattgaga acgatgaag atgaagaatg ttgttgagag gattgtgata      720

tttatatata cgaagattgg tttctggaga attcgatcat cttttttctcc attttcgtct      780

ctggaacgtt cttagagatg attgacgacg tgtcattatc tgatttgcag ttaaccaatg      840

ctttttgggt tggattcgtg gtacaccata ttatccgatt tggctcaatg gttttatata      900

aatttggttt tcggttcggt tatgagttat cattaaaatt aagctaacca aaaattttcg      960

taaaatttat ttcggtttca attcggatcc cttacttcca gaaccgaatt attcgaaacc     1020

ggggttagcc gaaccgaata ccaatgcctg attgactcgt tggctagaaa gatccaacgg     1080

tatacaataa tagaacataa atcggacggt catcaaagcc tcaaagagtg aacagtcaac     1140

aaaaaaagtt gagccctgag gagtatcgtt ccgccattt ctacgacgca aggcgaaaat     1200

ttttggcgcc aatctttccc ccctttcgaa ttctctcagc tcaaaacatc gtttctctct     1260

cactctctct cacaattcca aaaaatgcag cgccagagat cgattttgtc tttcttccaa     1320

aaacccacgg cggcgactac gaagggtttg gtttccggcg atgctgctag cggcgggggc     1380

ggcagcggag accacgattt aatgtgaagg aagggatgc taaaggcgac gcttctgtac     1440

gttttgctgt ttcgaaatct gtcgatgagg ttagaggaac ggatactcca ccggagaagg     1500

ttccgcgtcg tgtcctgccg tctggattta agccggctga atccgccggt gatgcttcgt     1560

ccctgttctc caatattatg cataagtttg taaaagtcga tgatcgagat tgttctggag     1620

agaggtacta atcttcgatt ctcttaattt tgttatcttt agctggaaga agaagattcg     1680
```

```
tgtaatttgt tgtattcgtt ggagagattc tgattactgc attggatcgt tgtttacaaa      1740

ttttcaggag ccgagaagat gttgttccgc tgaatgattc atctctatgt atgaaggcta      1800

atgatgttat tcctcaattt cgttccaata atggtaaaac tcaagaaaga aaccatgctt      1860

ttagtttcag tgggagagct gaacttagat cagtagaaga tataggagta gatggcgatg      1920

ttcctggtcc agaaacacca gggatgcgtc cacgtgcttc tcgcttgaag cgagttctgg      1980

aggatgaaat gacttttaag gaggataagg ttcctgtatt ggactctaac aaaaggctga      2040

aaatgctcca ggatccggtt tgtggagaga agaaagaagt aaacgaagga accaaatttg      2100

aatggcttga gtcttctcga atcagggatg ccaatagaag acgtcctgat gatccccttt      2160

acgatagaaa gaccttacac ataccacctg atgttttcaa gaaaatgtct gcatcacaaa      2220

agcaatattg gagtgttaag agtgaatata tggacattgt gcttttcttt aaagtggtta      2280

gtaactatta atctagtgtt caatccattt cctcaatgtg atttgttcac ttacatctgt      2340

ttacgttatg ctcttctcag gggaaatttt atgagctgta tgagctagat gcggaattag      2400

gtcacaagga gcttgactgg aagatgacca tgagtggtgt gggaaaatgc agacaggtaa      2460

attagttgaa acaactggcc tgcttgaatt attgtgtcta taaattttga caccaccttt      2520

tgtttcaggt tggtatctct gaaagtggga tagatgaggc agtgcaaaag ctattagctc      2580

gtgggtaagg gaaccatcat actttatgga attcgtttac tgctacttcg gctaggattt      2640

aagaaatgga aatcacttca agcatcatta gttaggatcc tgagaactca ggatgttttc      2700

ttattcgtta tataataagt cttttcatca aggagtaaca aacaaaactt gcacaatatt      2760

tgtgtgctca ctggcaaggc atatataccc agctaacctt tgctagttca ctgtagtaac      2820

agttacggat aatatatgtt tacttgtatg tggtaccctc attttgtctc tcatggaggc      2880

tttcaagcct tgtgttgaaa ctggatagtt acatatgctt ccaacagaaa ctagcatgca      2940

gattcatatg ctttcctatt ctactaatta tgtattgaca cactcgttgt ttcttttgaa      3000

agatataaag ttggacgaat cgagcagcta gaaacatctg accaagcaaa agccagaggt      3060

gctaatactg taagttttct tggataggtc aaggagagtg ttgcagactg tttttgatca      3120

tttctttttc tgtacattac tttcatgctg taattaactc aatggctatt ctggtctgat      3180

tatcagataa ttccaaggaa gctagttcag gtattaactc catcaacagc aagcgaggga      3240

aacatcgggc ctgatgccgt ccatcttctt gctataaaag aggtttgtta tttacttatt      3300
```

```
tatcttatca tgttcagttc atccaagtcc tgaaaaatta cactcttctt taccaatctt    3360

ccatcaagct gtgtaaagga tttggaatta gaaaatcatt atttgatgct ttgtttata     3420

tgcaagaggt tcccttgaaa agatctgttt aagattcttt gcacttgaaa aattcaatct    3480

ttttaagtga atcccctact ttcttacaat gatcatagtc tgcaattgca tgtcaagtaa    3540

tatcattcct tgttactgca tcccccctctt tcttaatgac cattgtctat gttgtgtttg   3600

tctcgtgtgc tggagaaaat gatagctgat ccaagctgta cattatcatg attaagtagc    3660

tgctcaggaa ttgcctttgg ttacattgcc taatggtttg atgtcaattt ttcttctgaa    3720

tctttatttt agatcaaaat ggagctacaa aagtgttcaa ctgtgtatgg atttgctttt    3780

gttgactgtg ctgccttgag gtttttgggtt gggtccatca gcgatgatgc atcatgtgct   3840

gctcttggag cgttattgat gcaggtaagc aagtgtattc tgtatcttat gtgtaccatg    3900

tgacttcctg tgcatatatt tgggttgcag gaactaattc tgaatcacca tttggtatgt    3960

ttttttccagg tttctccaaa ggaagtgtta tatgacagta aaggtaaact gcttgtatcg   4020

ccagttgttt tgttaaacag aatttaaggt aaatgacact ggttaattta aagtgcatac    4080

atgttgaaat attgcagggc tatcaagaga agcacaaaag gctctaagga aatatacgtt    4140

gacaggtacc atttcagtag gcaagctaac tgacaattta accgctcacc gaatgatagg    4200

tctcttaaac attgctaatg tagatgatgt ttatgtttca atctaatagg gtctacggcg    4260

gtacagttgg ctccagtacc acaagtaatg ggggatacag atgctgctgg agttagaaat    4320

ataatagaat ctaacggata ctttaaaggt tcttctgaat catggaactg tgctgttgat    4380

ggtctaaatg aatgtgatgt tgcccttagt gctcttggag agctaattaa tcatctgtct    4440

aggctaaagg tgtgttggct tgtttagttt ttgcttttca caaattaagc aaaggaactt    4500

ttcataactt acagtttcta tctacttgca gctagaagat gtacttaagc atggggatat    4560

ttttccatac caagtttaca ggggttgtct cagaattgat ggccagacga tggtaaatct    4620

tgagatattt aacaatagct gtgatggtgg tccttcaggc aagtgcatat ttcttttttg    4680

ataacttcaa ctagagggca gacatagaag gaaaaattct aatacttcgt acggatctcc    4740

agtaagtaat agccgatttt tgtttaccta tgtagggacc ttgtacaaat atcttgataa    4800

ctgtgttagt ccaactggta agcgactctt aaggaattgg atctgccatc cactcaaaga    4860

tgtagaaagc atcaataaac ggcttgatgt agttgaagaa ttcacggcaa actcagaaag    4920
```

```
tatgcaaatc actggccagt atctccacaa acttccagac ttagaaagac tgctcggacg      4980

catcaagtct agcgttcgat catcagcctc tgtgttgcct gctcttctgg ggaaaaaagt      5040

gctgaaacaa cgagtaagta tcaatcacaa gttttctgag taatgccttc catgagtagt      5100

ataggactaa aacattacgg gtctagctaa agactgttct ccttcttttg caatgtctgg      5160

ttattcatta catttctctt aacttattgc attgcaggtt aaagcatttg ggcaaattgt      5220

gaaagggttc agaagtggaa ttgatctgtt gttggctcta cagaaggaat caaatatgat      5280

gagtttgctt tataaactct gtaaacttcc tatattagta ggaaaaagcg ggctagagtt      5340

atttctttct caattcgaag cagccataga tagcgacttt ccaaattatc aggtgcccat      5400

ctatctttca tactttacaa caaaatgtct gtcactactc aaagcaatgc atatggctta      5460

gatctcaact cacaccccga ggatcctaaa gggatttgct ttttattcct aatgtttttg      5520

gatggtttga tttatttcta acttgaactt attaatcttg taccagaacc aagatgtgac      5580

agatgaaaac gctgaaactc tcacaatact tatcgaactt tttatcgaaa gagcaactca      5640

atggtctgag gtcattcaca ccataagctg cctagatgtc ctgagatctt ttgcaatcgc      5700

agcaagtctc tctgctggaa gcatggccag gcctgttatt tttcccgaat cagaagctac      5760

agatcagaat cagaaaacaa aagggccaat acttaaaatc caaggactat ggcatccatt      5820

tgcagttgca gccgatggtc aattgcctgt tccgaatgat atactccttg gcgaggctag      5880

aagaagcagt ggcagcattc atcctcggtc attgttactg acgggaccaa acatgggcgg      5940

aaaatcaact cttcttcgtg caacatgtct ggccgttatc tttgcccaag tttgtatact      6000

cgttagataa ttactctatt ctttgcaatc agttcttcaa catgaataat aaattctgtt      6060

ttctgtctgc agcttggctg ctacgtgccg tgtgagtctt gcgaaatctc cctcgtggat      6120

actatcttca caaggcttgg cgcatctgat agaatcatga caggagagag taagttttgt      6180

tctcaaaata ccaattcctc gaactattta ctcagatttt gtctgattgg acaaggtggt      6240

tttgcttttt tttaggtacc tttttggtag aatgcactga gacagcgtca gttcttcaga      6300

atgcaactca ggattcacta gtaatccttg acgaactggg cagaggaact agtactttcg      6360

atggatacgc cattgcatac tcggtaacct gctcttctcc ttcaacttat acttgttgat      6420

caacaaaaac atgcaattca ttttgctgaa acttattgat ttatatcagg tttttcgtca      6480

cctggtagag aaagttcaat gtcggatgct ctttgcaaca cattaccacc ctctcaccaa      6540
```

```
ggaattcgcg tctcacccac gtgtcacctc gaaacacatg gcttgcgcat tcaaatcaag    6600

atctgattat caaccacgtg gttgtgatca agacctagtg ttcttgtacc gtttaaccga    6660

gggagcttgt cctgagagct acggacttca agtggcactc atggctggaa taccaaacca    6720

agtggttgaa acagcatcag gtgctgctca agccatgaag agatcaattg gggaaaactt    6780

caagtcaagt gagctaagat ctgagttctc aagtctgcat gaagactggc tcaagtcatt    6840

ggtgggtatt tctcgagtcg cccacaacaa tgcccccatt ggcgaagatg actacgacac    6900

tttgttttgc ttatggcatg agatcaaatc ctcttactgt gttcccaaat aaatggctat    6960

gacataacac tatctgaagc tcgttaagtc ttttgcttct ctgatgttta ttcctcttaa    7020

aaaatgctta tatatcaaaa aattgtttcc tcgattataa caagattata tatgtatctg    7080

tcggtttagc tatggtatat aatatatgta tgttcatgag attggtcaag agaaatactc    7140

acaaacagta tattaagaag gaaatatgtt tatgcattaa tttaagtttc aagataaact    7200

gcaaataacc tcgactaaag ttgcaaagac caaacacaaa ttacaaaact tataagactt    7260

aagttctgaa ttccctaaaa ccaaaaaaaa aaacagaaca tattttgttg catctacaaa    7320

caacacaaac ctacatagtt tataacttac tcatcactga gattaacatc agaatcattc    7380

tccatttctt catcttcact ctcatcatca tcaccaccac catgatgatt ctcctcctct    7440

tcacgtaacc tagcaatctc actctgagct ctatcaacaa tctgcttctt ctgcaactcc    7500

aaatctctct gaaaatcagc tctcatcttc tccaactcct tcatttgctc tttcttactc    7560

ttctccatct tctcataaac cttcccaaac ctctcaacag aatccgccaa catcttatac    7620

gaagcagcgt cattaacctt cttcctctcg tactcaacct catcatcctc atcctcctcc    7680

tcttcagaat caccaggact atccatcatc tcatcaaacc cattagactt atctaaataa    7740

accttagtgt tcataaacac aaactcacct gaatcaacac cacaagctaa acctaaatcc    7800

gacttgggcg aaacacaaag caacatatcc aacttattga aaaacgacca tttacttgaa    7860

cctaaacctg atttctcaac cttaatcttc tcttttctat acttcctctt caagtcatca    7920

atcattctcc tacattgcgt ctcagatttc tccatcctta gctcctcact cactttctca    7980

gctacttcat tccaatcctc gttcctcaaa ctccttctac ccaattgcaa aaacctatct    8040

ccccaaactt caagcaacac aa                                            8062
```

PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : **C12N 15/82, 15/29, C07K 14/415, C12N 15/10, 5/04, 5/14, A01H 5/00**	A3	(11) International Publication Number: **WO 99/19492** (43) International Publication Date: 22 April 1999 (22.04.99)

(21) International Application Number: PCT/EP98/06977

(22) International Filing Date: 9 October 1998 (09.10.98)

(30) Priority Data:
PO 9745   10 October 1997 (10.10.97)   AU

(71) Applicant *(for all designated States except US)*: RHONE–POULENC AGRO [FR/FR]; 14/20, rue Pierre Baizet, F–69009 Lyon (FR).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DOUTRIAUX, Marie–Pascale [FR/FR]; 64, route de Villebon, F–91160 Saulx les Chartreux (FR). BETZNER, Andreas, Stefan [AU/AU]; 40 Dallachy Place, Page, ACT 2614 (AU). FREYSSINET, Georges [FR/FR]; 21, rue de Nervieux, F–69450 Saint Cyr au Mont d'Or (FR). PEREZ, Pascal [FR/FR]; 17, chemin de la Pradelle, Varennes, F–63450 Chanonat (FR).

(74) Agent: GENIN, Patrick; Rhône–Poulenc Agro, DPI, 14/20, rue Pierre Baizet, F–69009 Lyon (FR).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
24 June 1999 (24.06.99)

---

(54) Title: METHODS FOR OBTAINING PLANT VARIETIES

(57) Abstract

An isolated and purified DNA molecule comprising a polynucleotide sequence encoding a polypeptide functionally involved in the DNA mismatch repair system of a plant.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6   C12N15/82   C12N15/29   C07K14/415   C12N15/10   C12N5/04
        C12N5/14   A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6   C12N   A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 95 15381 A (CHAPELLE ALBERT DE ;UNIV JOHNS HOPKINS (US)) 8 June 1995 see the whole document ---	1-5, 33-35
A	WO 97 37011 A (SETRATECH S A R L ;BORTS RHONA HARRIET (GB); LOUIS EDWARD JOHN (GB) 9 October 1997 see abstract see the whole document ---	1-35
A	WO 90 07576 A (SETRATECH) 12 July 1990 see the whole document and specially page 5, line 16-31, examples 3-4 ---	1,7-27, 31,32

-/--

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

16 April 1999

Date of mailing of the international search report

03/05/1999

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Mateo Rosell, A.M.

Form PCT/ISA/210 (second sheet) (July 1992)

1

INTERNATIONAL SEARCH REPORT

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 01634 A (ANGELETTI P IST RICHERCHE BIO ;JIRICNY JOSEF (IT); PALOMBO FABIO ()) 16 January 1997 see page 1, line 18 - page 2, line 28 see page 58	1,2
A	PROLLA T A ET AL: "MLH1, PMS1, AND MSH2 INTERACTIONS DURING THE INITATION OF DNA MISMATCH REPAIR IN YEAST" SCIENCE, vol. 265, 19 August 1994, pages 1091-1093, XP000676403 cited in the application	1,2,4,5
A	A. WATANABE ET AL., : "Genomic organization and expresssion of the human MSH3 gene" GENOMICS, vol. 31, 1996, pages 311-318, XP002099967 see the whole document and specially Fig.3	1-6
A	ACHARYA S ET AL: "hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, no. 93, October 1996, pages 13629- 13629, XP002080159 cited in the application sequences from this paper deposited in GenBank under AC= U61981 (hMSH3), and U54777 and U73732-7 (hMSH6). see the whole document	1-6
A	K. LIU ET AL., : "characterization of the mouse Rep-3 gene: sequence similarities to bacterial and yeast mismatch-repair proteins" GENE, vol. 147, 1994, pages 169-177, XP002099968 see the whole document and specially Figure 2.	1-6
A	I. IACCARINO ET AL., : "MSH6, a Saccharomyces cerevisiae protein that binds to mismatches as a heterodimer with MSH2" CURRENT BIOLOGY, vol. 6, no. 4, April 1996, pages 484-486, XP002099969 sequence is deposited at GenBank under AC= Z47746. see the whole document	1-6

-/--

1

INTERNATIONAL SEARCH REPORT

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	CORRADI A ET AL: "CDNA SEQUENCE, MAP, AND EXPRESSION OF THE MURINE HOMOLOG OF GTBP, ADNA MISMATCH REPAIR GENE" GENOMICS, vol. 36, no. 2, 1 September 1996, pages 288-295, XP000613761 see the whole document and specially Fig.1	1-6
P,X	DATABASE EMBL NUCLEOTIDE AND PROTEIN SEQUENCES,13 July 1998, XP002099368 HINXTON, GB AC= AJ007791. Arabidopsis thaliana mRNA for mismatch repair protein (MSH3) see abstract	1-6
P,X	DATABASE EMBL NUCLEOTIDE AND PROTEIN SEQUENCES,1 August 1998, XP002099369 HINXTON, GB AC= O65607. Arabidopsis thaliana. Putative mismacht DNA repair protein see abstract	1-5
P,X	DATABASE EMBL NUCLEOTIDE AND PROTEIN SEQUENCES,16 October 1997, XP002099371 HINXTON, GB AC= AF009657. Arabidopsis thalian MutS homolog 6-like protein mRNA. see abstract -& K.M. CULLIGAN AND J.B. HAYS: "DNA mismatch repair in plants" PLANT PHYSIOLOGY, vol. 115, 1997, pages 833-839, XP002099372 see the whole document	1-6
T	DATABASE EMBL NUCLEOTIDE AND PROTEIN SEQUENCES,17 December 1998, XP002099373 HINXTON, GB AC= AJ131669. Triticum aestivum mRNA for MSH3 protein, MSH3 gene. see abstract	1-6
T	DATABASE EMBL NUCLEOTIDE AND PROTEIN SEQUENCES,12 October 1998, XP002099370 HINXTON, GB AC= AJ007792. Arabidopsis thaliana DNA mismatch repair protein, MSH6 gene. see abstract	1-6

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9515381	A	08-06-1995	EP	0730648 A	11-09-1996
			JP	9506509 T	30-06-1997
			US	5693470 A	02-12-1997
			US	5837443 A	17-11-1998
WO 9737011	A	09-10-1997	NONE		
WO 9007576	A	12-07-1990	FR	2641793 A	20-07-1990
			AT	127519 T	15-09-1995
			AU	4834390 A	01-08-1990
			CA	2006549 A	26-06-1990
			DE	68924174 D	12-10-1995
			DE	68924174 T	14-03-1996
			EP	0449923 A	09-10-1991
			ES	2077058 T	16-11-1995
			IE	72469 B	09-04-1997
			JP	4503601 T	02-07-1992
WO 9701634	A	16-01-1997	IT	RM950434 A	27-12-1996
			AU	6241296 A	30-01-1997

Form PCT/ISA/210 (patent family annex) (July 1992)